Modern Foundations of Light Transport Simulation

Christian Lessig

electromagnetic theory

configuration space Q

$$\frac{\partial}{\partial t}\begin{pmatrix}\vec{E}\\\vec{H}\end{pmatrix}=\begin{pmatrix}0 & -\frac{1}{\epsilon}\nabla\times\\\frac{1}{\mu}\nabla\times & 0\end{pmatrix}\begin{pmatrix}\vec{E}\\\vec{H}\end{pmatrix}$$

microlocal analysis
(Wigner transform)

phase space T^*Q

$$-\dot{W}^\varepsilon=\frac{1}{\varepsilon}[p^\varepsilon,W^\varepsilon]+\frac{1}{i}\{p^\varepsilon,W^\varepsilon\}+O(\varepsilon)$$

$\varepsilon\to 0$

$$\dot{W}_a^0+\{\tau_a,W_a^0\}=[W_a^0,F_a]$$

conservation of
frequency

unpolarized
radiation

Fermat's principle

$\hat{L}=n^2(q)$

Legendre
transform

light transport equation

$$\dot{\ell}=-\{\ell,H\}$$

cosphere
bundle reduction

classic iso-velocity description

$$S^*Q=(T^*Q\backslash\{0\})/\mathbb{R}^+$$

transport
theorem

measurements

ideal light transport:
globally defined
Hamiltonian vector field

classical radiometry

$$L(x,\omega)\cos\theta\,d\omega\,dA$$

Lie-Poisson structure of ideal light transport

$$\dot{\ell}=X_{\mathcal{H}_\ell}[\ell]=-\mathrm{ad}^*_{\frac{\delta\mathcal{H}_\ell}{\delta\ell}}(\ell)=-\{\ell,H\}$$

momentum
map

conservation of
ℓ along "rays"

Stone's theorem

$$\ell_t=\mathcal{U}_t\,\ell_0$$

$\mathcal{T}=\bar{\mathcal{U}}\,\mathcal{R}_\rho$

inclusion of
surface scattering

operator formulation of light transport

$$\bar{\ell}=\ell_0+\mathcal{T}^1\ell_0+\mathcal{T}^2\ell_0+\dots$$

finitary point functionals

reproducing
kernel bases

$$\ell(z)=\sum_{i=1}^{m}\ell(\lambda_i)\,\tilde{k}_i(z)$$

reproducing kernel Galerkin projection

$$P_k\bar{\ell}=P_k\ell_0+P_k\mathcal{T}^1P_k^{-1}P_k\ell_0+\dots$$

explicit

implicit

radiosity

path tracing
photon mapping

"Weil nicht Wirklichkeit wird,
 was man nicht vorher gedacht hat."

 Christa Wolf

Abstract

Light transport simulation aims at the numerical computation of the propagation of visible electromagnetic energy in macroscopic environments. In this thesis, we develop the foundations for a modern theory of light transport simulation, unveiling the geometric structure of the continuous theory and providing a formulation of computational techniques that furnishes remarkably efficacy with only local information.

To subsume light transport theory into modern mathematical physics, we derive its geometric formulation from Maxwell's equations from a representation on six dimensional phase space. At the short wavelength limit, unpolarized electromagnetic energy is then represented by the phase space light energy density whose propagation is governed by a Hamiltonian system with a light transport equation naturally afforded by the Poisson bracket. With a fiber-wise radial symmetry of the Hamiltonian, the dynamics on six dimensional phase space can be reduced to the five dimensional cosphere bundle, connecting our formulation to classical light transport theory. The connection to classical optics is obtained using a non-canonical Legendre transform which yields the Hamiltonian structure of light transport also from Fermat's principle, and which establishes that the light energy density propagates along a geodesic flow defined by the refractive index. Observables of the light energy density are obtained using the transport theorem of continuum mechanics, and an abstract description of measurements provides the connection to classical radiometry and quantities such as radiance and vector irradiance.

In idealized environments, light transport is a Lie-Poisson system for the group of symplectic diffeomorphisms, making it a natural phase space analogue of ideal fluid dynamics. The reduced Hamiltonian equation for ideal light transport is the light transport equation, describing coadjoint motion on the dual Lie algebra, and the Noetherian quantity is the light energy density, with the momentum map representing classical "radiance is constant along a ray".

The question our work on computational techniques for light transport simulation addresses is: how can efficient computations be performed with only the local information about the energy density that is available in the complex environments encountered in today's applications? Our answer is provided by reproducing kernel bases, practical Hilbert space representations whose expansion coefficients are given by function values. Founded on such representations, we introduce finitary point functionals, computational techniques that only employ local information from a countable set of locations, and which include many existing approaches for sampling, pointwise approximation, interpolation, and integration. With our formulation, close to optimal techniques are obtained by numerically optimizing point locations of a possibly overcomplete reproducing kernel basis, and this not only leads to more effective computations in applications but it also provides practical answers to long standing questions for example on numerical integration. By employing reproducing kernel bases for Galerkin projection, we obtain a Hilbert space formulation for sampling based techniques such as distribution ray tracing and path tracing, establishing a well defined interpretation with a finite number of samples and providing a unified perspective for a wide range of light transport simulation techniques, from radiosity over path tracing to photon mapping.

Founded on the belief that understanding guides the way to more effective computational techniques, this thesis presents the foundations for a modern theory of light transport simulation, a theory which still needs to be developed in its full elegancy and potential.

Contents

xiii

Acknowledgement

Many people shaped me and my thinking in the last years and enabled this thesis. In place for all, I would like to thank . . .

Chapter 1

Introduction

*"Wingless flying is cruel amusement
try it, clumsy, you'll fall over backwards."*[1]

In 1939, the state of radiometry was characterized as: "Theoretical photometry constitutes a case of 'arrested development', and has remained basically unchanged since 1760 while the rest of physics has swept triumphantly ahead. In recent years, however, the increasing needs [...] have made the absurdly antiquated concepts of traditional photometric theory more and more untenable."[2] The situation has not changed since then, and still Lambert's "antiquated" theory is employed. One purpose of the present thesis is to "catch up with the rest of mathematical physics" and to provide a description of light transport theory where it is subsumed into modern mathematical physics—its natural place as we will see. We will then be able to explain how light transport is a five and a half dimensional analogue of ideal fluid dynamics, and why Kelvin's circulation theorem describes the conservation of radiance along a ray. With modern mathematical foundations in hand, we will turn to the simulation of light transport and to the design of computational techniques that employ only finite local information—techniques that are effective with the pointwise values of the light energy density only available in applications. Our answer will be Hilbert space expansions whose coefficients are function values, and the use of such representations will be our *idée fixe* in the second part of the thesis. We will then be able to explain how to design low discrepancy sequences without paper and pencil, and why path tracing is a finite element method in a $50,331,648$-dimensional function space.

Our work on the present thesis was equally motivated by scientific curiosity and practical limitations, and it was guided by the belief that a thorough understanding provides the key to more effective computational techniques. Unfortunately, some parts of our understanding remain sketches at the moment.

---

[1] Daniil Kharms, from *The Aviation of Transformations*; translated by Matvei Yankelevich and Ilya Bernstein.

[2] From Moon and Timoshenko's preface to Gershun's book ("The Light Field").

Nonetheless, our thesis develops the foundations for a modern theory of light transport which explains how practical computational techniques provide approximations to Maxwell's equations. Filling in the sketches, and transforming the understanding into more effective computations will have to remain for the future.

Before we discuss the principles that guided our development and outline the work that will be developed in subsequent chapters, however, let us briefly consider the point of departure of the present thesis: light transport simulation in computer graphics and the shortcomings of the existing theory there.

1.1 Light Transport Simulation in Computer Graphics

In this section, we will summarize the history and understanding of light transport simulation in computer graphics. The shortcomings of the existing conception, which in our opinion impede progress towards more effective computational techniques, will be discussed thereafter.

1.1.1 A Historical Overview[3]

Early History The earliest work that aimed at a realistic depiction of synthetic environments was Appel's in the late 1960s.[4] Shortly afterwards, Goldstein and Nagel[5] followed a suggestion by Appel[6] and introduced ray tracing, and Kay[7] and Whitted[8] independently extended the work to account for the global transport effects that result from perfectly specular reflection and refraction. In the mid 1970s, Phong[9] developed an influential phenomenological shading model that incorporated specular reflection effects,[10] and Blinn currently pioneered the use of models from the optics[11] and radiative transfer literature,[12] pointing at the relevance of existing work in physics for computer graphics.[13] The results

---

[3]The technical material in this section can for example be found in (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*; Dutré, Bala, and Bekaert, *Advanced Global Illumination*).

[4]Appel, "The Notion of Quantitative Invisibility and the Machine Rendering of Solids"; Appel, "Some Techniques for Shading Machine Renderings of Solids".

[5]Goldstein and Nagel, "3-D Visual Simulation".

[6]Appel, "The Notion of Quantitative Invisibility and the Machine Rendering of Solids".

[7]Kay, "Transparency, Refraction, and Ray Tracing for Computer Synthesized Images".

[8]Whitted, "An Improved Illumination Model for Shaded Display".

[9]Phong, "Illumination for Computer-Generated Images"; Phong and Crow, "Improved Rendition of Polygonal Models of Curved Surfaces"; Phong, "Illumination for Computer Generated Pictures".

[10]Contrary to common belief, it was Warnock (*A Hidden Line Algorithm for Halftone Picture Representation*), and not Phong, who was the first to employ powers of the cosine term to obtain specular reflection effects.

[11]Blinn, "Models of Light Reflection for Computer Synthesized Pictures"; Torrance and Sparrow, "Theory for Off-Specular Reflection From Roughened Surfaces"; Trowbridge and Reitz, "Average Irregularity Representation of a Rough Surface for Ray Reflection".

[12]Blinn, "Light Reflection Functions for Simulation of Clouds and Dusty Surfaces"; Chandrasekhar, *Radiative Transfer*.

[13]Blinn, "Models of Light Reflection for Computer Synthesized Pictures".

by Phong and Blinn, and others such as Gouraud, Newell, Crow, and Catmull, laid the foundation for the work to come, and they established image synthesis as an independent research subject.

The Mathematization of Light Transport Simulation After Blinn's seminal application of results from the physics literature to model phenomena such as specular reflection or scattering from thin cloud layers, in the early 1980s Kajiya took up the study of physically based techniques for light transport simulation to obtain more general models. The efforts led to three seminal papers in which he employed results from the radiative transfer[14] and electromagnetic scattering theory literature[15] to introduce the volume transport equation,[16] a model for anisotropic reflection,[17] and the rendering equation[18] into the computer graphics literature. To this date, the rendering equation forms the mathematical basis of physically based image synthesis. It determines the steady state light flux density $\bar{L}(x, \bar{\omega})$ at a surface point x in direction $\bar{\omega}$, commonly known as radiance, as the sum of the emitted radiance $\tilde{L}(x, \bar{\omega})$ and the convolution integral of the incoming radiance $L(x, \omega)$ with the bidirectional reflection distribution function $\rho_x(\omega, \bar{\omega})$, that is

$$\bar{L}(x, \bar{\omega}) = \tilde{L}(x, \bar{\omega}) + \int_{H_x^2} L(x, \omega)\, \rho_x(\omega, \bar{\omega})\, |n_x \cdot \omega|\, d\omega, \tag{1.1}$$

where n_x is the surface normal, H_x^2 is the local hemisphere over x, and $|n_x \cdot \omega|$ is known as cosine factor or foreshadowing term, see Fig. 1.1. By the constancy of radiance along a ray in a non-scattering medium, incoming and outgoing radiance are related by

$$L(x, \omega) = \bar{L}(x', \bar{\omega}') = \bar{L}(t(x, \omega)) \tag{1.2}$$

where $t(x, \omega)$ is the ray tracing function which determines the closest surface point x' from x in direction ω. The bidirectional reflection distribution function is defined as

$$\rho_x(\omega, \bar{\omega}) = \frac{d\bar{L}(x, \bar{\omega})}{L(x, \omega)\, |n_x \cdot \omega|\, d\omega} \tag{1.3}$$

and it describes the directional distribution of scattered light at the scene location x. Using a change of variables from angle to area, the rendering equation can also be written as

$$\bar{L}(x_i \to x_{i+1}) = \tilde{L}(x_i \to x_{i+1})$$
$$+ \int_{\mathcal{M}} L(x_{i-1} \to x)\, \rho(x_{i-1} \to x_i \to x_{i+1})\, G(x_{i-1} \to x_i)\, dA(x_{i-1}) \tag{1.4}$$

---

[14]Chandrasekhar, *Radiative Transfer*.

[15]Beckmann and Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*.

[16]Kajiya and Herzen, "Ray Tracing Volume Densities".

[17]Kajiya, "Anisotropic Reflection Models".

[18]Kajiya, "The Rendering Equation".
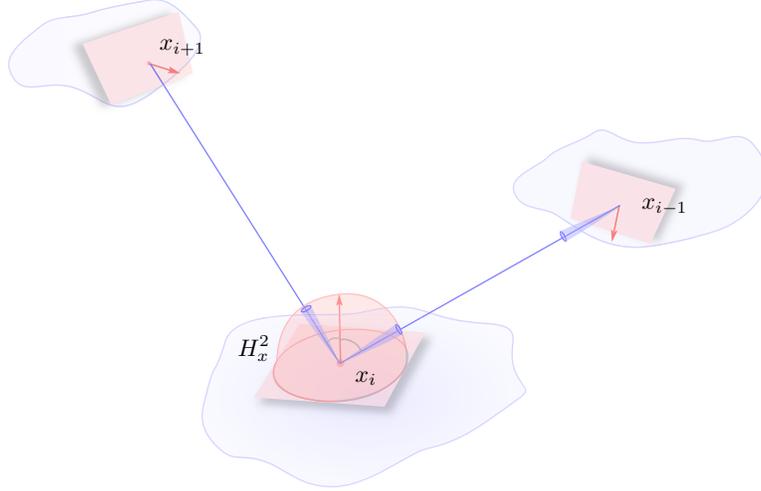
Figure 1.1: Geometry of light transport. Radiance $L(x, w)$ is defined for an infinitesimal cone around a surface direction ω at a surface point x. In the area formulation, the transport from x_{i-1} to x_{i+1} is considered with scattering at x_i.

where one now considers transport from a point x_{i-1} to a point x_{i+1} with scattering at x_i, and with the integration being over all scene surfaces $\mathcal{M}$, see again Fig. 1.1. The geometry term $G(x_{i-1} \to x_i)$ describes the dependence of the energy exchange on the scene geometry in Eq. 1.4 and it is given by

$$G(x_{i-1} \to x_i) = V(x_{i-1}, x_i) \frac{|n_{x-1} \cdot \omega| \, |n_x \cdot \omega|}{\|x_{i-1} - x_i\|^2} \tag{1.5}$$

where $V(x_{i-1}, x_i)$ is the binary visibility function, which in the angular formulation is implicitly in the ray tracing function. For convenience, we will in the following often write

$$\tilde{\rho}(x_{i-1} \to x_i \to x_{i+1}) = \rho(x_{i-1} \to x_i \to x_{i+1}) \, G(x_{i-1} \leftrightarrow x_i). \tag{1.6}$$

Together with the rendering equation, Kajiya also proposed an operator formulation for light transport. With a transport operator T defined as

$$T\bar{L} = \int_{S^2} \bar{L}(t(x, \omega)) \, \rho_x(\omega_o, \omega_i) \, |n \cdot \omega_i| \, d\omega, \tag{1.7}$$

and which combines the ray tracing function and the effect of the bidirectional reflection distribution function, Eq. 1.1 takes the form

$$\bar{L} = \tilde{L} + T\bar{L}. \tag{1.8}$$

By re-arranging terms one obtains

$$\tilde{L} = (I - T)L \tag{1.9}$$

where I is the identity operator. Since the operator norm $\|T\| < 1$ is strictly bounded by unity, which is satisfied due to energy conservation, the formal inverse of $(I - T)$ is given by the Neumann series. The solution of Eq. 1.8 is hence provided by

$$L = S\tilde{L} = \sum_{k=0}^{\infty} T^k \tilde{L} \qquad (1.10)$$

where S is known as solution operator, and the powers T^k of the transport operator T can be interpreted as light paths of length k.

The rendering equation models light transport in environments with scattering surfaces but where light propagates without disturbances in between. The volume transport equation, in contrast, describes the transport of radiance in an environment with scattering media. It is given by

$$\frac{\partial}{\partial t} \bar{L}(x, \bar{\omega}) = -\sigma_a(x, \bar{\omega}) \, L(x, -\bar{\omega}) + \sigma_s(x, \bar{\omega}) \int_{S^2} p(x, \bar{\omega}, \omega) \, L(x, \omega) \, d\omega \quad (1.11)$$

where x is now an arbitrary point in the scene, p is the phase function describing the angular distribution of scattered light, σ_a is the absorption coefficient accounting for the optical thickness of the medium, σ_s is the scattering probability, and we also assumed a medium which does not emit light.

Light Transport Simulation between Surfaces   Prior to the work by Kajiya which provided a theoretical basis for light transport simulation, algorithms yielding more realistic imagery than ray tracing emerged. Cook and co-workers introduced distribution ray tracing[19] where stochastic sampling is employed to include effects such as glossy reflection, soft shadows, and motion blur.[20] For the rendering equation in Eq. 1.1, for example, the technique employs the approximation

$$\bar{L}(x, \bar{\omega}) = \int_{S^2} L(x, \omega) \, \tilde{\rho}_x(\bar{\omega}, \omega) \, d\omega \approx \sum_{i=1}^{n} L(x, \omega_i) \, \tilde{\rho}_x(\bar{\omega}, \omega_i) \qquad (1.12)$$

where $\tilde{\rho}_x(\bar{\omega}, \omega) = \rho_x(\bar{\omega}, \omega) \, |n \cdot \omega|$ and the ω_i are random directions in the hemisphere above x. Shortly after the original work by Cook and co-workers, in the same paper where he introduced the rendering equation, Kajiya[21] showed that distribution ray tracing is a Monte Carlo approximation to the light transport problem. As an alternative solution technique but also based on the Monte Carlo principle, he proposed path tracing where the iterates $T^k \tilde{L}$ in the

---

[19]Cook, Porter, and Carpenter, "Distributed Ray Tracing"; Cook, "Stochastic Sampling in Computer Graphics".

[20]Random sampling had been used previously for the solution of light transport problems for example by Appel ("Some Techniques for Shading Machine Renderings of Solids"). Dippé and Wold ("Antialiasing through Stochastic Sampling") were other early proponents of stochastic techniques, although there motivation was to avoid aliasing.

[21]Kajiya, "The Rendering Equation".

Neumann series are estimated directly by averaging over paths $\mathbf{x}_i^k = (x_1^i, \ldots, x_k^i)$ of length k, that is

$$T^k \tilde{L} = \int_{\mathcal{M}^k} \tilde{L}(x_k \to x_{k-1}) \tilde{\rho}(x_k \to x_{k-1} \to x_{k-2}) \ldots$$

$$\ldots \tilde{\rho}(x_2 \to x_1 \to x_0) \, dA^k(\mathbf{x}_i^k) \qquad (1.13a)$$

$$\approx \sum_{i=1}^{n} \tilde{L}(x_k^i \to x_{k-1}^i) \, \tilde{\rho}(x_k^i \to x_{k-1}^i \to x_{k-2}^i) \, \ldots \, \tilde{\rho}(x_2^i \to x_1^i \to x_0^i). \qquad (1.13b)$$

The integral form of $T^k \tilde{L}$ in Eq. 1.13a is known as path integral formulation and in the computer graphics literature it first appeared in the work by Veach.[22] The formulation has also been employed to develop bidirectional path tracing[23] where T^k is split into $T^k = T^i T^j$, with $k = i + j$, and both T^i and the adjoint operator T^{j*} are employed for computations.[24] This amounts to starting paths from both the light sources and the camera, which is mathematically described by

$$I = \left\langle m, T^k \tilde{L} \right\rangle = \left\langle m, T^i T^j \tilde{L} \right\rangle = \left\langle T^{i*} m, T^j \tilde{L} \right\rangle \qquad (1.14)$$

where m is the measurement function, typically on the image plane, and I is a measurement, for example a pixel value. The Metropolis light transport algorithm, which was the first application of a Markov chain Monte Carlo method to the light transport problem, was also developed based on the path integral formulation.[25]

The use of Quasi Monte Carlo techniques for light transport simulation was pioneered by Shirley[26] and much work on the subject was done by Keller.[27] Although the benefits of the approach are still debated,[28] considerable practical improvements have been reported.[29]

In the 1990s, techniques for the solution of the rendering equation based on density estimation were developed for which a pointwise representation of the

---

[22]Veach, "Robust Monte Carlo Methods for Light Transport Simulation".

[23]Bidirectional path tracing was introduced independently by Veach and Guibas ("Bidirectional Estimators for Light Transport") and Lafortune and Willems (*Bi-Directional Path Tracing*; "A Theoretical Framework for Physically Based Rendering"). The most practical version so far was proposed by Kollig and Keller ("Efficient Bidirectional Path Tracing by Randomized Quasi-Monte Carlo Integration").

[24]In practice, the adjoint T^* has been employed implicitly since the early days of ray tracing. See the paper by Christensen for an account of adjoint methods for light transport simulation ("Adjoints and Importance in Rendering: an Overview").

[25]Veach and Guibas, "Metropolis Light Transport".

[26]Shirley, "Discrepancy as a Quality Measure for Sample Distributions".

[27]See for example (Keller, "Myths of Computer Graphics") and references therein.

[28]See for example (Shirley, Edwards, and Boulos, "Monte Carlo and Quasi-Monte Carlo Methods for Computer Graphics") and (Keller, "Myths of Computer Graphics").

[29]See for example (Kollig and Keller, "Efficient Bidirectional Path Tracing by Randomized Quasi-Monte Carlo Integration"; Kollig and Keller, "Efficient Bidirectional Path-Tracing by Randomized Quasi-Monte Carlo Integration"; Kollig and Keller, "Efficient Multidimensional Sampling"; Keller, "Myths of Computer Graphics"; Dammertz and Keller, "Image Synthesis by Rank-1 Lattices") and references therein.

outgoing radiance is employed. Although these techniques, such as irradiance caching[30] and photon mapping,[31] are only consistent Monte Carlo estimators,[32] they provide in practice usually considerably better performance than unbiased approaches.[33] The light cuts algorithm,[34] which was introduced recently, is another consistent technique that employs such a pointwise representation.

Concurrently to Cook's work on distribution ray tracing, Goral et al.[35] and Nishita and Nakamae[36] introduced the radiosity algorithm for perfectly diffuse environments,[37] adapting a classical technique from the heat transfer literature.[38] A theoretical justification for the algorithm was once again provided by Kajiya[39] who showed that it is an approximation to the rendering equation, and later work by Heckbert[40] explained that it is a finite element method obtained by Galerkin projection of the operator transport equation. Performing the projection for a projection operator P with left pseudo-inverse P^{-1} yields

$$PL = P\tilde{L} + P\left(TL\right) = P\tilde{L} + PTP^{-1}PL \tag{1.15}$$

which provides a finite approximation to the original equation with the matrix PTP^{-1} representing the transport operator. Later work on finite element methods investigated for example higher order radiosity methods,[41] extensions for glossy surfaces,[42] and stochastic radiosity techniques,[43] although these ideas

---

[30]Ward, Rubinstein, and Clear, "A ray tracing solution for diffuse interreflection"; Ward and Heckbert, "Irradiance Gradients"; Ward, "The RADIANCE lighting simulation and rendering system".

[31](Jensen and Christensen, "Photon Maps in Bidirectional Monte Carlo Ray Tracing of Complex Objects"; Jensen, "Importance Driven Path Tracing using the Photon Map"; Jensen, *Realistic Image Synthesis using Photon Mapping*); lesser known density estimation approaches are (Shirley et al., "Global Illumination via Density-Estimation"; Walter et al., "Global Illumination using Local Linear Density Estimation").

[32]In the literature, the techniques are often considered as biased, see for example (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*, Chapter 15.6), although it seems to us that it is rather apparent that they are consistent.

[33]A Monte Carlo integration technique is unbiased if the expected value of the estimator is the integral value for any number of samples, it is consistent if the integral value is obtained as the number of samples goes to infinity, and it is biased if even an infinite number of samples is not sufficient to obtain the correct solution.

[34]Walter et al., "Lightcuts: A Scalable Approach to Illumination"; Walter et al., "Multidimensional Lightcuts".

[35]Goral et al., "Modeling the Interaction of Light between Diffuse Surfaces".

[36]Nishita and Nakamae, "Calculation of Interreflections and Its Representation Method".

[37]The classic and still most comprehensive texts are (Sillion and Puech, *Radiosity and Global Illumination*) and (Cohen and Wallace, *Radiosity and Realistic Image Synthesis*).

[38]See for example the classic text by Howard and Siegel (*Thermal Radiation Heat Transfer*) for an introduction to heat transfer. An older but still relevant text is those by Planck (*Theorie der Wärmestrahlung*). The longer wavelength of heat radiation renders specular effects negligible.

[39]Kajiya, "The Rendering Equation".

[40]Heckbert and Winget, *Finite Element Methods for Global Illumination*; Heckbert, "Introduction to Finite Element Methods"; Heckbert, "Finite Element Methods for Radiosity".

[41]Zatz, "Galerkin Radiosity: A Higher Order Solution Method for Global Illumination"; Gortler et al., "Wavelet Radiosity"; Schröder et al., "Wavelet Projections for Radiosity".

[42]Christensen et al., "Global Illumination of Glossy Environments Using Wavelets and Importance"; Christensen et al., "Clustering for Glossy Global Illumination".

[43]Neumann et al., "The Stochastic Ray Method for Radiosity"; Feda, "A Monte Carlo

received only limited attention and have rarely been considered since the 1990s. More recently, precomputed radiance transfer[44] was introduced to obtain limited global transport effect at real-time frame rates, and it was subsequently shown that the technique is a radiosity-like approximation when the illumination is assumed to be independent of the scene location.[45]

Light Transport Simulation in Volumes Computational techniques for volume transport have received only limited attention in the past,[46] and most existing algorithms are adaptations of techniques for surface transport for volumetric effects, such as finite element methods,[47] photon mapping,[48] Metropolis light transport,[49], and bidirectional path tracing.[50] The high complexity of the volume transport problem makes thereby usually approximations such as the single scattering assumption or the diffusion approximation necessary.[51]

Next to volume scattering, a medium can affect the transport of radiance also through a varying refractive index. In the computer graphics literature, however, no transport equation for this phenomenon exists, and instead the eikonal equation from geometric optics is employed and heuristically combined with the required concepts from radiometry. Correspondingly, the problem has received only limited attention in the past.[52]

1.1.2 Limitations of the Current Conception

In the foregoing, we provided an outline of the development and understanding of light transport simulation in computer graphics, and since the 1960s a rapid evolution from Appel's first work on the subject to today's complex simulations could be observed. In recent years, however, efforts have concentrated on

---

Approach for Galerkin Radiosity"; Sbert, "The Use of Global Random Directions to Compute Radiosity – Global Monte Carlo Techniques"; Sbert, "Error and Complexity of Random Walk Monte Carlo Radiosity".

[44]Sloan, Kautz, and Snyder, "Precomputed Radiance Transfer for Real-Time Rendering in Dynamic, Low-Frequency Lighting Environments"; Kautz, Sloan, and Snyder, "Fast, Arbitrary BRDF Shading for Low-Frequency Lighting using Spherical Harmonics".

[45]Lehtinen, "A Framework for Precomputed and Captured Light Transport".

[46]For a survey see (Cerezo et al., "A survey on participating media rendering techniques").

[47]Rushmeier and Torrance, "The Zonal Method for Calculating Light Intensities in the Presence of a Participating Medium"; Rushmeier, "Realistic Image Synthesis for Scenes with Radiatively Participating Media".

[48]Jensen and Christensen, "Efficient simulation of light transport in scences with participating media using photon maps".

[49]Pauly, Kollig, and Keller, "Metropolis Light Transport for Participating Media".

[50]Lafortune and Willems, "Rendering participating media with bidirectional path tracing".

[51]The diffusion approximation was introduced in (Kajiya and Herzen, "Ray Tracing Volume Densities") but it was fully developed only in (Stam, "Multiple scattering as a diffusion process").

[52]Berger, Trout, and Levit, "Ray tracing mirages"; Musgrave and Berger, "A note on ray tracing mirages (comments and author's reply)"; Gröller, "Nonlinear Ray Tracing: Visualizing Strange Worlds"; Stam and Languénou, "Ray Tracing in Non-Constant Media"; Gutierrez et al., "Simulation of atmospheric phenomena"; Ihrke et al., "Eikonal Rendering: Efficient Light Transport in Refractive Objects"; Zhao et al., "Visual simulation of heat shimmering and mirage."

incremental improvements to existing techniques. This situation is easily attributed to a maturing field. It is our belief, however, that the slowdown is caused by the current theoretical foundations, and that it is their shortcomings which hamper the development of existing and alternative research directions. Although a judgement is difficult with no alternatives at hand, we will discuss some of these limitations in the following, and the reader is invited to reflect on them again at the end of the thesis with our formulation in hand.

In contrast to other theories such as ideal fluid dynamics[53] or elasticity[54] whose mathematical foundations were revisited over time, light transport theory has not received such a reformulation. Hence, still Lambert's "antiquated concepts"[55] are in use, and even more recent developments such as the transport equation correspond no longer to the mathematics common elsewhere in science and engineering. One could argue that the time-honoured formulation served well and still fulfils its purpose. However, with concepts such as radiance—which were developed even 100 years before vector calculus became widely used—many ideas and problems cannot be formulated, and no true understanding of their mathematical and physical significance is possible. Practically, this for example impedes the development of techniques that incrementally update an image when a scene description is slightly perturbed, since even when visibility is ignored the infinitesimal changes caused by moving objects currently cannot be formulated mathematically.[56] Similarly, at the moment no mathematically well founded description of the transport of radiance derivative values is possible, preventing for instance the development of sampling based techniques that use derivative information to better amortize the high computational costs of obtaining a sample.

After the initial efforts by Kajiya who explained how light transport simulation can be formulated as a finite element and Monte Carlo method, little principal work on the foundations of computational techniques exists in the computer graphics literature. In fact, not even the three central questions which today impede progress towards more effective computations have been clearly identified:

1. How can visibility be determined efficiently?

2. How can the "curse of dimensionality" be broken?

3. How can effective computations be performed with only local information?

---

[53]The last reinvention of the theory of fluid dynamics began in the 1960s, cf. (Arnold, "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits"; Arnold, "Sur un principe variationnel pour les découlements stationaires des liquides parfaits et ses applications aux problèmes de stabilité non linéaires"; Ebin and Marsden, "Groups of Diffeomorphisms and the Motion of an Incompressible Fluid"; Khesin, "Topological Fluid Dynamics"; Arnold and Khesin, *Topological Methods in Hydrodynamics*).

[54]See for example (Marsden and Hughes, *Mathematical Foundations of Elasticity*; Kanso et al., "On the geometric character of stress in continuum mechanics") and references therein.

[55]From the preface by Moon and Timoshenko to Gershun's book ("The Light Field").

[56]A singular exception is the work by Chen and Arvo ("Theory and Application of Specular Path Perturbation") where the idea has been explored.

Without being widely known, these questions have also only rarely addressed,[57] arguably also because of their difficulty. Additionally, other important question have not been considered to this date. For example, in computer graphics no conceptualization of an image exists that reconciles the continuous models of the mathematical theory with the finite descriptions unavoidable for computations,[58] preventing for instance that sampling on the image plane is performed in a function space that is better adapted to natural images than the "pixel domain".[59] Another unanswered question is which function spaces naturally arise for light transport and which spaces should be considered for computations. This for instance prevents to adaptation of computational techniques for these settings, which gravely limits their efficiency. For example, when generic Monte Carlo techniques are employed, the known local smoothness of the light energy density cannot be exploited, and, additionally, these techniques are not well defined for a finite number of samples and not computational tractable in the functional analytic settings currently assumed.[60] The trichotomy of computational techniques that are employed in practice—Monte Carlo algorithms, finite element methods, and density estimation techniques—is similarly unsatisfactory, making it for instance difficult to compare the approaches, and explain the significantly superior performance of density estimation in practice.

1.2 Towards Modern Foundations of Light Transport

As we showed in the last section, the existing theory on light transport simulation in computer graphics led to an impressive progress in the field in the past 40 years.[61] However, we also showed that the current foundations leave many questions open and prevent the development of many promising research directions. We therefore believe that advancing the understanding of the mathematical foundations of light transport simulation is not only of scientific interest

---

[57]A notable exception is (Dachsbacher et al., "Implicit Visibility and Antiradiance for Interactive Global Illumination") and (Durand, Drettakis, and Puech, "The 3D Visibility Complex: A New Approach to the Problems of Accurate Visibility"; Durand, Drettakis, and Puech, "The Visibility Skeleton: A Powerful and Efficient Multi-Purpose Global Visibility Tool"; Durand, Drettakis, and Puech, "The 3D Visibility Complex").

[58]Image models were discussed before in the graphics literature, see for example (Fiume, *The Mathematical Structure of Raster Graphics*; Smith, *A Pixel Is Not a Little Square*; Blinn, "What Is a Pixel?") but this work leaves many conceptual questions open, and it does not agree with the powerful ideas which are employed in other fields such as signal processing, cf. (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*).

[59]The rare exceptions we are aware of are (Bolin and Meyer, "A Frequency Based Ray Tracer"; Overbeck, Donner, and Ramamoorthi, "Adaptive Wavelet Rendering").

[60]Arvo ("Analytic Methods for Simulated Light Transport") and Veach ("Robust Monte Carlo Methods for Light Transport Simulation"), the only researchers who considered functional analytic questions, assumed the Lebesgue space L_2. However, computations such as the approximation of an integral require in this setting an exponential number of samples to attain an ϵ-approximation, making them computationally intractable. We will return to this question in later chapters but already refer to (Traub and Werschulz, *Complexity and Information*).

[61]Jensen and Akenine-Möller, "The Race for Real-Time Photorealism".

but also vital for the development of more effective computational techniques.[62]

In the following, we will first consider the question how modern foundations can be obtained and what constitutes an effective programme towards this objective, and subsequently we will outline the theory which will be developed in detail in Chapter 3 and Chapter 4.

1.2.1 Towards where?

The limitations of the current foundations of light transport simulation that were outlined pose the question how these can be overcome? In the past, Arvo and Veach tried to provide answers to this question, and later Durand and Ramamoorthi considered theoretical aspects of light transport simulation.[63] Arvo, for example, showed that mathematical foundations for the volume transport equation can be obtained from classical transport theory, and he derived the equation using the particle assumptions employed there.[64] However, as will be discussed in more detail in the sequel, abstract transport theory does not provide a physical basis for light transport theory, and with the mathematics employed by Arvo only Lambert's "antiquated concepts" are recovered. Together with Veach, Arvo also suggested the use of measure theory to overcome the limitations of the current mathematical formulation of radiometry.[65] The generality of measures and the lacking notion of smoothness renders them however ineffective to model physical phenomena and to develop efficient computational techniques. An analysis of the functional analytic structure of light transport was also initiated by Arvo and Veach, but their work remained in its infancy and provided few practical benefits. Recently, Durand and collaborators[66] and Ramamoorthi and co-workers[67] studied light transport using techniques from classical signal processing. However, these tools are not suited for the nontrivial manifolds encountered in light transport,

---

[62]Our work hence follows the motto put forth by Simo, Marsden, and Krishnaprasad for elasticity: *"It is our belief that a thorough understanding of the mathematical underpinnings of elasticity is crucial to its analytical and numerical implementation."*, ("The Hamiltonian structure of nonlinear elasticity: The material and convective representations of solids, rods, and plates", p. 1).

[63]Arvo, "Transfer Equations in Global Illumination"; Arvo, "The Role of Functional Analysis in Global Illumination"; Arvo, "Analytic Methods for Simulated Light Transport"; Veach, "Robust Monte Carlo Methods for Light Transport Simulation".

[64]Classical texts on the subject are for example (Wing, *An Introduction to Transport Theory*; Spanier and Gelbard, *Monte Carlo Principles and Neutron Transport Problems*; Duderstadt and Martin, *Transport Theory*).

[65]The work builds upon and was inspired by those by Preisendorfer (*Radiative Transfer on Discrete Spaces*).

[66]Durand et al., "A Frequency Analysis of Light Transport"; Soler et al., "Fourier Depth of Field"; Egan et al., "Frequency Analysis and Sheared Reconstruction for Rendering Motion Blur".

[67]Ramamoorthi and Hanrahan, "On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object"; Ramamoorthi and Hanrahan, "A Signal-Processing Framework for Reflection"; Ramamoorthi, Mahajan, and Belhumeur, "A First-Order Analysis of Lighting, Shading, and Shadows"; Mahajan et al., "A Theory of Locally Low Dimensional Light Transport"; Mahajan, Tseng, and Ramamoorthi, "An Analysis of the In-Out BRDF Factorization for View-Dependent Relighting".

and they were only employed based on the existing formulation, which renders them ineffective to address its principal limitations.

The inadequacy of work in the literature leaves us with the quest to establish our own programme to improve the understanding of the mathematical and computational foundations of light transport simulation. But even developing such a programme leaves us with many questions:

What characterizes a better understanding?
What determines a suitable mathematical language?
What is the relationship between continuous and computational models?
What determines an appropriate model for simulations?
What establishes the effectiveness of a simulation technique?
What characterizes efficient techniques for light transport simulation?

Characterizing what affords a better understanding is challenging—and will in part always remain subjective. However, the utility of a theory provides guidance. For example, scientifically a formulation which elucidates a system's structure and reveals connections to other physical phenomena is of intrinsic value, while computationally the ability to predict and explain experimental results is of utility. When simulations are effective is applications dependent. For us, it should provide qualitatively correct prediction,[68] and it should be possible to understand the error that arises through unavoidable approximations. The effectiveness of computations depends crucially on the choice of an appropriate setting—in mathematics, a result is established in the most general form possible, but for computations it is vital to exploit the natural structure of a problem, and almost surely a technique designed for a different or more general setting will not provide an answer most effectively.

With the above considerations in mind, we developed the mathematical and computational foundations of light transport theory. We employed geometric mechanics, the language of much of modern mathematical physics, to elucidate the structure and symmetries of light transport, and the geometric formulation is expressed using tensor analysis and exterior calculus, mathematical tools that naturally operate on the six dimensional phase space on which light transport is defined, and where classical vector calculus, which serves so well for systems in three dimensional space, is no longer available. The use of geometric mechanics is not only vital to expose the structure of light transport, but it also lays the foundation for the development of structure preserving integrators: numerical techniques that provide discrete analogues of continuous systems, and thereby yield qualitatively correct simulations independent of the degrees of freedom employed for the computations. Our geometrization of light transport will arise from Maxwell's equations in a medium with varying refractive index,

---

[68]The importance of obtaining qualitatively accurate predictions is particularly apparent for the simulation of systems where no or little experimental validation is possible, a classical example being simulations to determine the stability of the solar system (Laskar, "Large-scale chaos in the solar system"; Tremaine, "Is the solar system stable?"; Laskar and Gastineau, "Existence of collisional trajectories of Mercury, Mars and Venus with the Earth.").

Maxwell's equations ensuring that our work is well founded in modern physics, and a varying refractive index revealing the structure which is obscured in a homogeneous medium—as the shape of a three dimensional object can often not be discerned from a two dimensional projection. Using our geometric formulation, we will employ representation theory to obtain an alternative, functional analytic description of light transport, and this will enable us to develop an operator formulation paralleling those in the classical literature. With this bridge to existing work, we will reconsider the foundations of computational techniques and provide an answer to one of the three central question that were raised in the foregoing in Chapter 1.1.2: what provides a formulation of light transport simulation that is mathematically and practically tractable but restricted to finite local information, restrained to the pointwise values of the energy density that are only available? The principal setting of our answer will be familiar Hilbert spaces, where basis expansions enable to efficiently work with continuous functions, but where we will differ from previous work is the use of reproducing kernels as basis functions, which yields function values as expansion coefficients. Moreover, for such reproducing kernel bases close to optimal computational techniques are obtained by numerically optimizing the point locations and by employing overcomplete representations with more samples than necessary, and the Hilbert space setting also enables to analyze the error that arises in applications. Extending this idea, and using reproducing kernel bases together with Galerkin projection enables us to recover sampling based techniques, such as distribution ray tracing, path tracing, and photon mapping, within a Hilbert space setting. This provides insight into their working principles and ensuring a well defined interpretation with a finite number of samples. Additionally, reproducing kernel Galerkin projection yields a common mathematical framework for a wide range of computational techniques for light transport simulation.

1.2.2 Modern Foundations of Light Transport Simulation

After discussing our conceptual approach, we will in the following provide an overview of the work which will be developed in Chapter 3 and Chapter 4. A graphical depiction can be found on the following page. The diagram is also provided on page ii at the beginning of the thesis, and one should consult it whenever details overwhelm the structure of our arguments.

The Geometry of Light Transport   Light transport describes the propagation of visible electromagnetic energy in macroscopic environments. To subsume the theory into modern mathematical physics, we will found our development in Maxwell's equations on $Q \subset \mathbb{R}^3$ for a medium with possibly varying refractive index $n : Q \subset \mathbb{R}^3 \to \mathbb{R}$. Instead of considering the transport of the energy density $\mathcal{E}(q)$ at the short wavelength limit on configuration space $Q \subset \mathbb{R}^3$, however, we follow the literature and employ microlocal and semi-classical analysis to lift Maxwell's equations to phase space T^*Q. There, the electromagnetic field is represented by a 6×6 symmetric tensor density W^ε, the Wigner transform of

electromagnetic theory

configuration space Q

$$\frac{\partial}{\partial t}\begin{pmatrix} \vec{E} \\ \vec{H} \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{\epsilon}\nabla\times \\ \frac{1}{\mu}\nabla\times & 0 \end{pmatrix}\begin{pmatrix} \vec{E} \\ \vec{H} \end{pmatrix}$$

microlocal analysis
(Wigner transform)

phase space T^*Q

$$-\dot{W}^\varepsilon = \frac{1}{\varepsilon}\left[p^\varepsilon, W^\varepsilon\right] + \frac{1}{i}\{p^\varepsilon, W^\varepsilon\} + O(\varepsilon)$$

$\varepsilon \to 0$

$$\dot{W}_a^0 + \{\tau_a, W_a^0\} = \left[W_a^0, F_a\right]$$

unpolarized
radiation

conservation of
frequency

Fermat's principle

$$\hat{L} = n^2(q)$$

Legendre
transform

light transport equation

$$\dot{\ell} = -\{\ell, H\}$$

cosphere
bundle reduction

classic iso-velocity description

$$S^*Q = (T^*Q\setminus\{0\})/\mathbb{R}^+$$

transport
theorem

measurements

classical radiometry

$$L(x, \omega)\,\cos\theta\,d\omega\,dA$$

ideal light transport:
globally defined
Hamiltonian vector field

Lie-Poisson structure of ideal light transport

$$\dot{\ell} = X_{\mathcal{H}_\ell}[\ell] = -\mathrm{ad}^*_{\frac{\delta\mathcal{H}_\ell}{\delta\ell}}(\ell) = -\{\ell, H\}$$

momentum
map

conservation of
ℓ along "rays"

Stone's theorem

$$\ell_t = \mathcal{U}_t\,\ell_0$$

$$\mathcal{T} = \bar{\mathcal{U}}\,\mathcal{R}_\rho$$

inclusion of
surface scattering

operator formulation of light transport

$$\bar{\ell} = \ell_0 + \mathcal{T}^1\ell_0 + \mathcal{T}^2\ell_0 + \ldots$$

finitary point functionals

reproducing
kernel bases

$$\ell(z) = \sum_{i=1}^{m} \ell(\lambda_i)\,\tilde{k}_i(z)$$

reproducing kernel Galerkin projection

$$P_k\bar{\ell} = P_k\ell_0 + P_k\mathcal{T}^1 P_k^{-1} P_k\ell_0 + \ldots$$

explicit

implicit

radiosity

path tracing
photon mapping

the electromagnetic vector fields $(\vec{E}, \vec{H})$. For unpolarized radiation the tensor density W^ε becomes in the short wavelength limit $\varepsilon \to 0$ a scalar density

$$\ell = \ell(q,p) = \mathcal{L}(q,p)\,dq\,dp \in \mathrm{Den}(T^*Q) \tag{1.16}$$

which we denote as phase space light energy density since the fiber integral

$$\mathcal{E}(q) = \|\vec{E}\|^2 + \|\vec{H}\|^2 = \int_{T_q^*Q} \ell \tag{1.17}$$

recovers the electromagnetic energy density $\mathcal{E}(q)$ of the visible radiation $(\vec{E}, \vec{H})$. At the short wavelength limit, the dynamics on phase space T^*Q are governed by a canonical Hamiltonian system which determines a flow $\eta : \mathbb{R} \times T^*Q \to T^*Q$. The generator of the flow is the Hamiltonian

$$H(q,p) = \frac{c}{n(q)}\|p\| \tag{1.18}$$

and by Hamilton's equations it determines a vector field $X_H = (\dot{q}, \dot{p}) \in T(T^*Q)$ given by

$$\frac{\partial q}{\partial t} = \frac{\partial H}{\partial p} = \frac{c}{n(q)}\frac{p}{\|p\|} \tag{1.19a}$$

$$-\frac{\partial p}{\partial t} = \frac{\partial H}{\partial q} = c\,\|p\|\frac{\nabla_q n(q)}{n^2(q)} \tag{1.19b}$$

which provides an infinitesimal description of the flow. The time evolution of the light energy density $\ell \in \mathrm{Den}(T^*Q)$ is determined by the light transport equation

$$\dot{\ell} = -\{\ell, H\} = -\mathscr{L}_{X_H}\ell. \tag{1.20}$$

which is given by the negative canonical Poisson bracket $\{\,,\}$ on phase space, or, equivalently, by the negative Lie derivative $\mathscr{L}_{X_H}\ell$ along the Hamiltonian vector field X_H. The Hamiltonian structure of light transport can also be obtained from Fermat's principle using a non-canonical Legendre transform, and this shows that the propagation on phase space describes a geodesic flow along a metric

$$g_{ij}^n = \begin{pmatrix} n^2(q)/c^2 & 0 & 0 \\ 0 & n^2(q)/c^2 & 0 \\ 0 & 0 & n^2(q)/c^2 \end{pmatrix} \tag{1.21}$$

defined by the refractive index $n : \mathbb{R}^3 \to \mathbb{R}$. The Hamiltonian for light transport satisfies

$$H(q, \gamma p) = \gamma H(q,p) \ , \ \gamma \in \mathbb{R}^+ \tag{1.22}$$

and it is hence positive homogeneous of degree one. This enables to reduce the flow $\eta^\theta : \mathbb{R} \times T^*Q\backslash\{0\} \to T^*Q\backslash\{0\}$ on the slit cotangent bundle $T^*Q\backslash\{0\}$ to a flow $\tilde{\eta} : \mathbb{R} \times S^*Q \to S^*Q$ on the cosphere bundle

$$S^*Q = (T^*Q\backslash\{0\})/\mathbb{R}^+ \,, \tag{1.23}$$

the quotient of $T^*Q \backslash \{0\}$ by the multiplicative action $m_\gamma : (q, p) \mapsto (q, \gamma p)$ of $\mathbb{R}^+$. The reduced dynamics are then described by

$$
\begin{array}{ccc}
T^*Q \backslash \{0\} & \xrightarrow{\ \eta_t^\theta\ } & T^*Q \backslash \{0\} \\[4pt]
m_\gamma \uparrow & & \uparrow m_\gamma \\[4pt]
S^*Q & \xrightarrow[\ \tilde{\eta}_t\ ]{} & S^*Q
\end{array}
\tag{1.24}
$$

although for the light energy density, no such reduction is easily possible since the $\mathbb{R}^+$ fibers contain the frequency dependence of $\ell \in \mathrm{Den}(T^*Q)$, information which is vital in most applications. Nonetheless, for a fixed frequency, the description of the dynamics on the cosphere bundle S^*Q provides the connection to the five dimensional space of positions and directions that is classically employed to describe light transport.

Measurements of the light energy density can be obtained using the transport theorem, a central pillar of continuum mechanics. This provides a rigorous justification for the cosine term, omnipresent in classical light transport theory, and it bridges the gap to radiometry, with concepts such as radiance and vector irradiance arising as the surface independent parts of measurements.

The Group Structure of Ideal Light Transport When the Hamiltonian vector field for light transport is defined globally, the theory affords an elegant description as a Lie-Poisson system whose configuration space is the group $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ of symplectic diffeomorphisms. For the field Hamiltonian

$$
\mathcal{H}_\ell = \mathcal{H}(\ell_t) = \int H(q, p)\, \ell_t(q, p)\, \varpi = \int \frac{c}{n(q)}\, \|p\|\, \ell_t(q, p)\, \varpi,
\tag{1.25}
$$

the light transport equation is the reduced Hamiltonian equation in the Eulerian representation

$$
\dot{\ell} = X_{\mathcal{H}_\ell}[\ell] = -\mathrm{ad}^*_{\frac{\delta \mathcal{H}_\ell}{\delta \ell}}(\ell) = -\left\{ \ell, \frac{\delta \mathcal{H}_\ell}{\delta \ell} \right\} = -\{\ell, H\}
\tag{1.26}
$$

and it is the infinitesimal generator of the coadjoint action of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ on the dual Lie algebra $\mathfrak{g}^* \in \mathrm{Den}(T^*Q)$ whose elements are light energy density fields $\ell \in \mathrm{Den}(T^*Q) \cong \mathfrak{g}^*$. The momentum map for light transport yields the convective light energy density as Noetherian quantity that is conserved along the flow of the system. An explicit calculation shows that the momentum map is equivalent to the classical law of "conservation of radiance along a ray", and that it is the phase space analogue of Kelvin's circulation theorem for the conservation of convective vorticity. Next to the momentum map, additional conserved quantities exist in the form of enstrophy integrals. The Lie-Poisson structure of light transport for a globally defined Hamiltonian vector field establishes a remarkable similarity to ideal fluid dynamics, enabling to consider

the systems as configuration and phase space analogues of each other and motivating the name 'ideal light transport'.

In the Hamiltonian formulation of light transport, the light energy density is transported along a vector field. Alternatively, by interpreting the Hamiltonian vector field as an anti-self-adjoint operator acting on functions, Stone's theorem allows to introduce a real unitary transport operator

$$\mathcal{U}(t) \,:\, \mathbb{R} \times \mathcal{H}(T^*Q) \to \mathcal{H}(T^*Q) \tag{1.27}$$

which describes light transport as a flow on the space $\mathcal{H}(T^*Q)$ of phase space light energy densities. Using $\mathcal{U}(t)$, a surface transport operator $\bar{\mathcal{U}} : \mathcal{H}(S^+\mathcal{M}) \to \mathcal{H}(S^-\mathcal{M})$ can be defined which maps from outgoing to incoming light energy density on the scene surfaces $\mathcal{M}$. Together with a scattering operator $\mathcal{R}_\rho : \mathcal{H}(S^-\mathcal{M}) \to \mathcal{H}(S^+\mathcal{M})$, which describes the interaction at surfaces and maps from incoming to outgoing energy density, this yields a scattering transport operator

$$\mathcal{T} = \bar{\mathcal{U}} \, \mathcal{R}_\rho \,:\, \mathcal{H}(S^+\mathcal{M}) \to \mathcal{H}(S^+\mathcal{M}), \tag{1.28}$$

the analogue of the transport operator in the classical literature. Denoting the emitted light energy density as $\ell_0 \in \mathcal{H}(S^+\mathcal{M})$, one then recovers the existing formulation of the steady state light energy density $\bar{\ell} \in \mathrm{Den}(T^*Q)$ as the superposition

$$\bar{\ell} = \ell_0 + \mathcal{T}^1\ell_0 + \mathcal{T}^2\ell_0 + \ldots = \sum_{i=0}^{\infty} \mathcal{T}^i\ell_0 \tag{1.29}$$

of the iterates $\ell^k = \mathcal{T}^k\ell_0$ of the scattering transport operator.

Foundations of Computational Techniques After developing a geometric theory of light transport, we address how efficient numerical simulations are possible: what provides a formulation of computational techniques that is mathematically and practically tractable but employs only the local information about the light energy density that is available in applications? A suitable setting to work computationally with continuous functions are Hilbert spaces $(H(X), \langle\,,\rangle)$ where bases $\{\phi_i\}_{i=1}^n$ allow the representation of functions as

$$f(x) = \sum_{i=1}^{n} \langle f(y), \phi_i(y) \rangle \, \phi_i(x) = \sum_{i=1}^{n} f_i \, \phi_i(x) \tag{1.30}$$

and where the scalar basis function coefficients f_i can be employed for numerical computations. However, computing the coefficients f_i requires the evaluation of inner products $f_i = \langle f(y), \phi_i(y) \rangle$, usually by integration, while the only information that is available are function values $f(\lambda_i)$ at a discrete set of locations $\lambda_i \in X$. We circumvent the computation of inner products—while preserving the advantages of a Hilbert space representation—through the use of reproducing kernel functions $k_y(x)$ that satisfy

$$f(y) = \langle k_y(x), f(x) \rangle \tag{1.31}$$

and which have the reproducing property of the point evaluation functional $\delta_y(x)$ but are elements in the Hilbert space $\mathcal{H}(X)$. The reproducing kernel enables the construction of possibly overcomplete "basis" expansions where functions $k_i(x) = k_{\lambda_i}(x)$ "anchored" at a set of reproducing points $\Lambda = \{\lambda_i\}_{i=1}^m$ are employed as basis functions. With the associated dual kernel functions being $\tilde{k}_i(x)$, the representation for an arbitrary function $f \in \mathcal{H}(X)$ is given by

$$f(x) = \sum_{i=1}^m \langle f(y), k_i(y) \rangle \, \tilde{k}_i(x) = \sum_{i=1}^m f(\lambda_i) \, \tilde{k}_i(x). \tag{1.32}$$

The above expansion in a reproducing kernel basis provides our answer on how to efficiently work with signals when only local information is available: the use of a basis expansions ensures mathematical and practical tractability, and the $k_i(x)$ yield basis coefficients that are given by function values instead of inner products. Moreover, for finite dimensional Hilbert spaces—the setting necessarily encountered for practical computations—the reproducing kernel is

$$k_y(x) = k(y, x) = \sum_{i=1}^n \phi_i(y) \, \phi_i(x) \tag{1.33}$$

which enables the construction of reproducing kernel bases for $\mathcal{H}(X)$ by choosing a suitable set of locations $\Lambda = \{\lambda_i\}_{i=1}^m$, a choice which will typically also determine the properties of the representation. A numerical representation for a reproducing kernel basis is given by

$$\mathrm{K}_\phi(\Lambda) = \begin{bmatrix} \phi_1(\lambda_1) & \cdots & \phi_n(\lambda_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\lambda_m) & \cdots & \phi_n(\lambda_m) \end{bmatrix} \in \mathbb{R}^{m \times n} \tag{1.34}$$

and we call $\mathrm{K}_\phi(\Lambda)$ the reproducing matrix with respect to $\{\phi_i\}_{i=1}^n$. The matrix $\mathrm{K}_\phi(\Lambda)$ provides the change of basis from $\{\phi_i\}_{i=1}^n$ to the reproducing kernel basis $\{k_i\}_{i=1}^m$. Hence the basis function coefficients $f(\phi) = \{f_i(\phi)\}$ with respect to $\{\phi_i\}_{i=1}^n$ can be reconstructed from the point values $f(\Lambda) = \{f(\lambda_i)\}$, or basis function coefficients with respect to $\{k_i(x)\}_{i=1}^m$, as

$$f(\phi) = \mathrm{K}_\phi(\Lambda)^{-1} f(\Lambda). \tag{1.35}$$

This provides a practical means to obtain the $f_i(\phi)$ when only local information is available. The change of basis with the inverse kernel matrix also suggests to employ the condition number $\mathrm{cond}(\mathrm{K}_\phi(\Lambda))$, which determines the accuracy with which the above linear system can be solved, as a quality measure for a reproducing kernel basis. We can then define tight and nearly tight reproducing kernel bases, optimal and nearly optimal representations with respect to the condition number, which are overcomplete analogues of orthonormal bases. In contrast to the common situation in the literature, such representations can be constructed efficiently by numerically optimizing the point locations λ_i, avoiding the laborious efforts traditionally required to find "good" sampling points. For many applications, such numerically tight or nearly tight reproducing kernel

bases yield nearly optimal computational techniques, with the robustness of the representations being controlled by the overcompleteness of the "basis", which is vital for example when the properties of the input signal cannot be controlled or are not known.

Founded on these ideas, we introduce finitary point functionals: computational techniques that employ only local information from a countable set of locations, and whose foundation is naturally provided by reproducing kernel bases. As finitary point functionals, sampling theorems and pointwise approximation schemes are at once given when the elements of a function space are represented in a reproducing kernel basis, and the dual kernel functions then provide a rigorous interpretation of classical reconstruction filters. Similarly, interpolation techniques are obtained with reproducing kernel bases which are not overcomplete and where the dual kernel functions naturally satisfy

$$\langle k_i(x), \tilde{k}_j(x) \rangle = \tilde{k}_j(\lambda_i) = \delta_{ij} \tag{1.36}$$

and are hence interpolatory. By linearity, integration techniques can be formulated as finitary point functional as

$$\int_X f(x)\, dx = \int_X \sum_{i=1}^{m} f(\lambda_i)\, \tilde{k}_i(x)\, dx \tag{1.37a}$$

and exploiting linearity yields

$$\int_X f(x)\, dx = \sum_{i=1}^{m} f(\lambda_i) \int_X \tilde{k}_i(x)\, dx. \tag{1.37b}$$

By defining weights w_i as

$$w_i = \int_X \tilde{k}_i(x)\, dx \tag{1.37c}$$

we obtain the quadrature rule

$$\int_X f(x)\, dx = \sum_{i=1}^{m} w_i\, f(\lambda_i). \tag{1.37d}$$

Following the literature and in close analogy to the classical result for Quasi Monte Carlo integration, we show that the error of the above quadrature rule can be characterized by a generalized Koksma-Hlawka inequality.

Finitary point functionals are designed for a suitable function space. However, in many applications it cannot be guaranteed that the signal is contained in the space. We analyze the error that arises when a residual signal component outside of the space exists, known as aliasing error in classical nomenclature, and we bound it by the magnitude of the residual and a term which depends on the reproducing points $\Lambda = \{\lambda_i\}$ that are employed.

As examples for finitary point functionals we derive Monte Carlo integration, Gauss-Legendre quadrature, Lagrange interpolation, and the Shannon sampling

theorem from the general theory by making suitable choices for the function space and the reproducing points. Monte Carlo integration, for instance, can be obtained by considering the space spanned by characteristic or indicator functions $\{\chi_i\}_{i=1}^n$ defined over a partition of the integration domain X. When the χ_i all have the same size $|\chi_i|$, then the weights in Eq. 1.37c become $w_i = |X|/n$, which immediately yields the standard Monte Carlo estimator for uniform sampling by Eq. 1.37d. Similarly, when the size of the partitions is determined by a function $p : X \to \mathbb{R}$ and no longer uniform, the Monte Carlo estimator for importance sampling is recovered.

To demonstrate the potential of finitary point functionals, we consider two applications: the rotation of finite spherical harmonics expansions and the fiber-wise projection of monochromatic light energy density into a basis. For the rotation of signals represented in spherical harmonics, we construct a sampling theorem for the sphere and exploit that the action of the rotation group $SO(3)$ on functions is pointwise by the pullback $(\mathrm{R}f)(\bar{\omega}) = f(\mathrm{R}^{-1}\bar{\omega})$ for $\bar{\omega} \in S^2$. By linearity, a rotated signal can then be obtained by reconstruction from rotated sampling points

$$\mathrm{R}f_l = \sum_{i=1}^{m} \langle \mathrm{R}f_l(\omega), k_i(\omega) \rangle \, \tilde{k}_i(\omega) = \sum_{i=1}^{m} f_l(\mathrm{R}^{-1}\lambda_i) \, \tilde{k}_i(\omega) \qquad (1.38)$$

and by altering the location and number of the locations λ_i we obtain techniques that are well adapted for different applications. For the fiber-wise basis projection of monochromatic light energy density, we consider the space $\mathcal{H}_{\leq L}(S^2)$ spanned by all spherical harmonics up to band L as approximation space, and construct overcomplete and numerically optimized reproducing kernel bases for the setting. The sought after spherical harmonics coefficients can then be obtained with the inverse kernel matrix $\mathrm{K}_\phi(\Lambda)^{-1}$ and Eq. 1.35. In practice, the input signals are never perfectly contained in $\mathcal{H}_{\leq L}(S^2)$. Nonetheless, our experimental results demonstrate that our ansatz is considerably more efficient than Monte Carlo and Quasi Monte Carlo integration. As our error analysis shows, the key to this efficiency is, firstly, an approximation space that is better adapted to the properties of the input signals than those considered by generic techniques, and, secondly, the use of optimized sampling locations and oversampling to increase the robustness of the technique to aliasing error arising from residual signal components.

Reproducing Kernel Galerkin Projection   Existing techniques for light transport simulation can be categorized as finite element, Monte Carlo, or density estimation methods, a trichotomy which makes it for example difficult to compare different techniques and explain their different performance characteristics. Using Galerkin projection, one of the "workhorse's" of computational science and engineering, we provide a formulation of light transport simulation where a wide range of techniques, including those based on sampling and density estimation, have a common functional analytic formulation. The key to this common perspective is Galerkin projection using reproducing kernel bases,

which provides a functional analytic interpretation of samples. For example, for a first order integral equation

$$g(y) = \int_X f(x)\rho(x,y)\,dx, \tag{1.39a}$$

such as the shading equation, the ansatz yields by representing the input $f(x)$ in a tight reproducing kernel basis that

$$g(y) = \int_X \sum_{i=1}^m f(\lambda_i)\,k_i(x)\,\rho(x,y)\,dx. \tag{1.39b}$$

Exploiting linearity one obtains

$$g(y) = \sum_{i=1}^m f(\lambda_i)\int_X k_i(x)\,\rho(x,y)\,dx \tag{1.39c}$$

and with the reproducing property of the $k_i(x)$ one has

$$g(y) = \sum_{i=1}^m f(\lambda_i)\,\rho(\lambda_i,y). \tag{1.39d}$$

Projecting also the outgoing signal into the reproducing kernel basis yields

$$\langle g(y), k_j(y)\rangle = \left\langle \sum_{i=1}^m f(\lambda_i)\,\rho(\lambda_i,y),\, k_j(y)\right\rangle \tag{1.39e}$$

and by once again using the reproducing property of the kernel functions one obtains

$$g(\lambda_j) = \sum_{i=1}^m f(\lambda_i)\,\rho(\lambda_i,\lambda_j). \tag{1.39f}$$

The last equation has the form of an estimator that only employs point values. However, in contrast to classical interpretations of Monte Carlo and density estimation techniques, the Hilbert space formulation provides insight into computations with a finite number of samples, for example under which conditions a numerically exact solution is obtained, and it enables a characterization of the error. Moreover, algorithms such as radiosity, distribution ray tracing, path tracing, and photon mapping differ from this perspective only in the basis and the parametrization that is employed.

Our derivation relies on the existence of a tight or nearly tight reproducing kernel basis. Unfortunately, we are currently not able to precisely characterize when such representations exists, and our formulation hence remains formal in this respect. Nonetheless, one class of function spaces where the derivations can be carried out are those spanned by characteristic basis functions, which connects these results for light transport simulation to our general observation for Monte Carlo integration.

1.2.3   A Guide to the Reader

A central subject of the present thesis is an understanding of light transport simulation not available with existing formulations. Unsurprisingly, this required much effort on finding effective languages for such an endeavour, and lead us to an amalgam of ideas from mathematics, computer science, and physics that locates the present thesis at the intersection of the fields. Having consequently material from disciplines with widely different lines of thought and cultures, choices have to be made on how it is presented and what can be included, and still most readers will probably be disappointed: mathematicians by the lack of rigour, computer scientists by the lack of applications, and physicists by the lack of previously unknown physics.

With a thesis spanning multiple disciplines and fields, one also faces the question for whom it should be written, and which expository material needs to be included. Although we tried to be as inclusive as space and time permitted, the "role model" of a reader we had in mind most of the time was a computer scientist, a point of departure which made it necessary to include extensive mathematical background material to develop the concepts employed in the remainder of the thesis which are usually not part of a computer science curriculum. However, unless these contribute to the understanding of the material, we omit all proofs, which was also necessary to not further increase the already extensive length of the background section.

We begin in Chapter 2.2 with the required notions of functional analysis, material which is today fairly standard in many areas of computational science and engineering. However, in Chapter 2.2.3 our treatment also include reproducing kernels Hilbert spaces, which are central to our own work but often neglected in the literature. Readers who are usually not concerned with numerical questions should also refer to Chapter 2.2.4 for how computations with continuous signals are possible on a finite computer. Central to much of geometric mechanics is calculus on manifolds, and we will in fact require a treatment which goes beyond what is often found in introductory books on the subject. Such a treatment will be developed in Chapter 2.3.2. The presentation there is best suited as a reference to refer back to it when needed, and the reader might only want to have a casual look at the material at a first reading. In Chapter 2.3.4, we will develop geometric mechanics, and in particular the Hamiltonian point of view of the theory. Our presentation aims thereby at a more intuitive perspective than can be found in the literature, and will we hence include more worked example than is common. Mathematically the most challenging material is arguably Chapter 2.3.5 where the description of mechanical systems using Lie groups is discussed. We again aiming at a presentation which is more intuitive than is common in the literature, but the reader is cautioned that a thorough understanding requires in our own experience considerable effort.

The reader might wonder at times in the following if *all* the presented background material is necessary for our thesis. Indeed, some concepts will only serve as stepping stones, and others will only be casually related to our developments. However, we believe that stepping stones are often vital to truly

understand a concept, and that a broad exposition is required to avoid that ideas remain "islands in a terra incognita". Additionally, the present thesis aims at the foundations for a modern theory of light transport simulation, and hence we did not hesitate to include material where we believe it will be helpful in the future to complete the programme.

An extensive index has been included, and the reader should employ it extensively to refer back to definitions and earlier discussions as needed. Additionally, a list of symbols and a flow diagram of the thesis can be found at the end of the thesis.

Chapter 2

Mathematical Foundations

"Nun gut wer bist du denn? Ein Teil von jener Kraft,
die stets das Böse will und stets das Gute schafft."[1]

In this chapter, the mathematical language employed in the remainder of the thesis will be introduced. We will thereby assume originality modest mathematical background, and emphasize intuition over mathematical formalism. Proofs will thus be omitted unless insight into the mathematics is provided.

We will begin by recalling some essential notions from algebra in Chapter 2.1. Chapter 2.2 contains a discussion of linear vector spaces and the concepts from functional analysis required for Chapter 4. In Chapter 2.3, we will develop geometric mechanics, beginning from the foundations provided by the calculus on manifolds to the most elegant and challenging aspects of the theory such as the description of a mechanical system as a flow on a Lie group. Although we cannot entirely avoid set theory and topology, for background on this material we have to refer to the literature.[2]

2.1 Preliminaries

In the following, it will be useful to have in mind some notions from algebra about the relationship between mathematical structures. We will therefore recall them at this point.

Definition 2.1. *A **homomorphism** is a structure preserving map between algebraic structures.*

Homomorphisms exists for different structures, such as algebras, Lie algebras, rings, and groups, and the precise meaning of the term depends on the structure of interest. The following examples demonstrate this.

---

[1] Johann Wolfgang von Goethe, *Faust: Der Tragödie erster Teil*, 1808.

[2] For set theory see for example (Bourbaki, *Elements of Mathematics: Theory of Sets*), and an accessible introduction to the essential notions of topology be found in (Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*, Chapter 1).

Example 2.1. Let $(\mathcal{H}, \langle , \rangle)$ be a finite dimensional function space that is Hilbert space, and let $\{\varphi_i\}$ and orthonormal basis for $\mathcal{H}$. Then $\{\varphi_i\}$ defines a Hilbert space homomorphism from $\mathcal{H}$ to $\mathbb{R}^n$ with the usual Euclidean inner product.

Example 2.2. A Lie algebra homomorphism $\varphi : (\mathfrak{g}, [,]_\mathfrak{g}) \to (\mathfrak{h}, [,]_\mathfrak{h})$ between two Lie algebras $(\mathfrak{g}, [,]_\mathfrak{g})$ and $(\mathfrak{h}, [,]_\mathfrak{h})$ is a map such that $[X, Y]_\mathfrak{g} = [\varphi(X), \varphi(Y)]_\mathfrak{h}$ for all $X, Y \in \mathfrak{g}$.

Definition 2.2. *An **isomorphism** between two algebraic structures is a homomorphism whose inverse is also a homomorphism.*

An isomorphism thus defines an equivalence between two structures that enables to perform all operations on the first structure also with the second one, and then relate the result back to the first. When two structures A and B are isomorphic then this will be denoted as $A \cong B$.

Example 2.3. Extending Example 2.1, any finite dimensional Hilbert space $(\mathcal{H}, \langle , \rangle)$ of dimension n is isomorphic to Euclidean space $(\mathbb{R}^n, \cdot)$ with the inner product over $\mathcal{H}$ being the dot product over $\mathbb{R}^n$.

Definition 2.3. *An **endomorphism** is a homomorphism from an object onto itself.*

Definition 2.4. *An **automorphism** is an isomorphism from an object onto itself.*

Automorphisms are for example of importance in the context of (Lie) groups when it is acting on itself.

Definition 2.5. *A **homeomorphism** $\varphi : S \to T$ between topological spaces S and T is a bijective mapping that is continuous and has a continuous inverse.*

By the required continuity, a homeomorphism is an isomorphism in the category of topological spaces, preserving the topological structure of S and T. Some care is required to distinguish homomorphisms and homeomorphisms which are related—and orthographically almost coincide—but nonetheless distinct concepts.

2.2 Applied Functional Analysis[3]

Much of the world around us is best described as being continuous—a description not only of much intuitive appeal but also of astonishing mathematical power.

---

[3]The material in this section is largely drawn from (Lax, *Functional Analysis*; Christensen, *An Introduction to Frames and Riesz Bases*) and the reader is referred there or for example to (Rudin, *Functional Analysis*; Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*) for more detailed treatments. (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*) is a standard reference on more applied aspects. References are largely omitted in this section, although no claim of originality of the presented material is made whatsoever.

In this section, we will develop a language to reason about continuous objects and to perform computations with them on a finite computer. The principal setting for this is provided by linear spaces formed by continuous functions, signals describing physical phenomena, and the additional structure which becomes available when for example a norm or an inner product is introduced. Our treatment will not shy away from infinite dimensional spaces when these are necessary or convenient, but we will also never forget that finite dimensional settings are unavoidable for computations.[4] Importantly, the linear theory of the present section will also furnish the foundation—and the blueprint—for the nonlinear calculus on manifolds that will be developed in Chapter 2.3.2, and which will provide the fabric for the development of geometric mechanics in subsequent sections.

We will begin in Chapter 2.2.1.1 by introducing linear spaces and studying the structure provided by them, and the natural mappings between such spaces will be considered in Chapter 2.2.1.2. In Chapter 2.2.1.3, introducing a norm, a length measure for linear spaces, will enable us consider Banach spaces, while Hilbert spaces will be obtained in Chapter 2.2.2.1 when we equip a linear space with an inner product. Operators on Hilbert spaces will be discussed in Chapter 2.2.2.2, and we will introduce the central concepts of a basis and a frame in Chapter 2.2.2.3 and Chapter 2.2.2.4, respectively. In Chapter 2.2.3, we will consider reproducing kernel Hilbert space which will provide the setting for much of Chapter 4, and computations with continuous signals on a finite computer will be discussed in Chapter 2.2.4. We conclude the section by developing three examples in detail in Chapter 2.2.5.

One might ask why we begin with linear space and develop the theory "ground up", when all we are interested in are computations in finite function space? However, by developing the mathematical context of our setting we will obtain a perspective that enables an understanding of how our constructions differ from whose obtained in the literature, and why we are able to provide simple answers to questions which are seemingly perplexing elsewhere. Additionally, the more general perspective will also prove very useful when we develop manifold theory in Chapter 2.3.2.

2.2.1 Linear Spaces

In this section, we will develop the elementary theory of linear spaces, or vector spaces, and important associated concepts such as linear maps.

2.2.1.1 Foundations

Linear spaces are modelled on groups and we will hence briefly recall the definition of a group.

---

[4]This is the spirit found in much of the literature on modern signal processing, cf. (ibid., Chapter 1).

Definition 2.6. *A **group** G is a set with a binary **group multiplication***

$$g \circ h = gh : G \times G \to G, \quad g, h \in G$$

*that is closed in the set and associative so that $f \circ (g \circ h) = (f \circ g) \circ h$. The **identity element** e of a group is the unique element such that*

$$e \circ g = g \circ e = g$$

*for all $g \in G$. For every group element $g \in G$ there exists a unique **inverse element** g^{-1} such that*

$$g \circ g^{-1} = g^{-1} \circ g = e.$$

*A group is **Abelian** if group multiplication commutes and $g \circ h = h \circ g$ for all $g, h \in G$.*

Groups will play an important role in the following when we discuss geometric mechanics, but for the moment the above definition suffices for our purposes. We are now prepared to introduce the central object of interest in the present section.

Definition 2.7. *A **linear space** or **vector space** V over the real numbers $\mathbb{R}$ is a set together with two binary operations:*

  i) addition of elements x, y in V,

$$x + y : V \times V \to V;$$

  ii) multiplication by scalars a in $\mathbb{R}$,

$$ax : \mathbb{R} \times V \to V.$$

The set V forms an Abelian group with respect to addition '+' with the identity element 0 so that

$$x + 0 = x,$$

and the inverse of each element $x \in V$ is denoted by $-x$ satisfying

$$x + (-x) = x - x = 0.$$

Multiplication by scalars a, b is assumed to be associative so that

$$a(bx) = (ab)x$$

and distributive for elements so that

$$a(x + y) = ax + by.$$

Linear space can be finite and infinite dimensional, and they can be discrete or continuous. We will clarify these notions in the following, but the distinction should be kept in mind and plays an important role in subsequent chapters.

Example 2.4. Euclidean space $\mathbb{R}^n$ is a vector space with the usual addition of vectors and multiplication by scalars.

Historically, Euclidean space served as model for the concept of a linear space, which took its modern form in the hands of Banach and Hilbert. In the following, it will be useful to keep the example of $\mathbb{R}^n$ in mind, which always allows to obtain concrete representations for the concepts that are introduced.

Example 2.5. The space of polynomials over the real line with addition being defined pointwise and the usual scalar multiplication forms a vector space.

Remark 2.1. Linear spaces can be defined over arbitrary fields F. We will confine ourselves to the case $F = \mathbb{R}$, which is the setting relevant for the applications we have in mind.

Next, we will introduce composite linear spaces consisting of two parts.

Definition 2.8. *The **direct sum** V$\oplus$W of two linear space V and W consists of order pairs (v, w) for $v \in$ V and $w \in$ W, with addition and scalar multiplication defined component-wise.*

The usual Euclidean space $\mathbb{R}^n$ is hence the direct n-sum of the real line $\mathbb{R}$.

Definition 2.9. *A **linear subspace** U $\subset V$ of a linear space V is a subset of V that is closed under addition and scalar multiplication.*

For example, Euclidean space $\mathbb{R}^k \subset \mathbb{R}^n$ is a linear subspace of $\mathbb{R}^n$ for $k \leq n$. Important properties of linear subspaces are summarized in the following proposition.

Proposition 2.1. *Let* V *be a linear space. Then*

 i) *$\{0\}$ and V are linear subspaces;*

 ii) *the sum of any collection of linear subspaces is a linear subspace;*

iii) *the intersection of any collection of subspaces is a linear subspace;*

Another important result concerning linear subspaces is the following.

Proposition 2.2. *Let* V *be a linear space and* U *a linear subspace of* V*. An **orthogonal complement of** U is the linear subspace* W *of* V *such that*

$$V = U \oplus W.$$

Next, we introduce the linear subspace which "fills" an arbitrary subset of a linear space V.

Definition 2.10. *Let S be a subset of the linear space* V. *The **linear span** $\mathrm{span}\,(S) \subset \mathrm{V}$ **of** S is the intersection of all linear subspace of* V *containing S,*

$$\mathrm{span}\,(S) = \left\{ \bigcap \mathrm{U} \mid \mathrm{U}\ \text{linear subspace of } V \text{ containing } S \right\}.$$

An alternative characterization of the linear span of a subset $S \subset V$ is provided in the following proposition.

Proposition 2.3. *Let* V *be a linear space and $S \subset \mathrm{V}$ be a subset. Then*

i) $\mathrm{span}\,(S)$ *is the smallest linear subspace of* V *containing S;*

ii) $\mathrm{span}\,(S)$ *consists of all **linear combinations** $y \in \mathrm{V}$ of the form*

$$y = \sum_{i}^{n} a_i\, x_i$$

where $x_i \in S$, $a_i \in \mathbb{R}$, and n any finite natural number.

The second part of the above proposition is of much importance since it allows to characterize $S \subset V$ by "representative elements" x_i. The next definition introduces a basic notion of "minimality" for linear combinations that will prove useful in the following.

Definition 2.11. *Let* V *be a linear space, and let $\{x_i\}_{i \in I}$ be a set of elements $x_i \in V$ for some index set I. A linear combination $\sum_{i \in I} a_i\, x_i$ is **trivial** if $a_i = 0$ for all i in the sum. The set $\{x_i\}_{i \in I}$ is **linearly independent** when*

$$0 = \sum_{i}^{n} a_i\, x_i$$

*implies that $a_i = 0$ for all i, that is when all non-trivial linear combinations of elements $x_i \in X$ are nonzero, and $\{x_i\}_{i \in I}$ is said to be **maximally independent** when the set ceases to be linearly independent when any nontrivial element in* V *is added to it. The set $\{x_i\}_{i \in I}$ is **complete in** V *when*

$$\mathrm{span}_{i \in I}\,(x_i) = \mathrm{V}.$$

2.2.1.2   Linear Maps

Next, we will introduce natural maps for linear space, which, unsurprisingly, are maps preserving the two elementary operations defined for these spaces.

Definition 2.12. *Let* V *and* W *be linear spaces. Then a mapping $T : \mathrm{V} \to \mathrm{W}$ is a **linear map** when*

$$T(x + y) = T(x) + T(y)$$
$$T(a\,x) = a\,T(x)$$

*for $x, y \in V$ and $a \in \mathbb{R}$. The **domain** of T is* V, *and its **target** is* W.

From an algebraic point of view, a linear map is hence a homomorphism in the category of linear spaces. Often, one also counters the case of an endomorphism, that is that T is a map $T : V \to V$ from V onto itself, cf. Def. 2.3.

Example 2.6. Let V be a linear space. Then the identity mapping $I : V \to V$ defined by $Ix = x$ for $x \in V$ is a linear map.

Example 2.7. Let A be the vector space of infinite sequences

$$(\ldots, a_{-k}, \ldots, a_0, \ldots, a_l, \ldots)$$

with $a_i \in \mathbb{R}$ and vector space addition defined element-wise. Then the left shift operator $L : A \to A$ defined by $La_i = a_{i+1}$ is a linear map.

When a linear map is bijective and surjective then it is an isomorphism of linear spaces, cf. Def. 2.2. We can exploit this to define the dimension of a linear space.

Definition 2.13. *Let* V *be a linear space. The* ***dimension*** $\dim(V)$ ***of*** V ***is*** n *when it is isomorphic to* $\mathbb{R}^n$ *for* $n < \infty$. *Otherwise, we say* V *is* ***infinite dimensional*** *and write* $\dim(V) = \infty$.

An isomorphism between linear spaces also leads to the following notion.

Definition 2.14. *A linear mapping* $T : V \to W$ *is* ***invertible*** *when it is bijective and surjective onto* W. *T then has an* ***inverse*** T^{-1} *satisfying*

$$T^{-1}T = I \qquad T\,T^{-1} = I$$

where I *is the identity mapping on the respective linear spaces.*

An important property of the inverse is the following.

Proposition 2.4. *Let* $T : U \to W$ *be an invertible linear map. Then the inverse is a linear map* $T^{-1} : W \to U$.

The inverse as defined in Def. 2.14 satisfies both $T^{-1}T = I$ and $TT^{-1} = I$. When only one of the conditions is satisfied one obtains the following maps.

Definition 2.15. *Let* $T : V \to W$ *be a linear map. A* ***left pseudo-inverse*** T_L^{-1} ***of*** T *is a linear map* $T_L^{-1} : W \to V$ *such that*

$$T_L^{-1}T = I,$$

where I *is the identity mapping on* V, *and a* ***right pseudo-inverse*** T_R^{-1} ***of*** T *is a linear map* $T_R^{-1} : W \to V$ *such that*

$$T\,T_R^{-1} = I$$

where I *is the identity mapping on* W.[5]

---

[5]In the literature, one sometimes also finds 'generalized inverse' instead of 'pseudo-inverse'.

Note that a left or right pseudo-inverse is usually not unique, and different choices can provide different advantages and disadvantages in applications. In practice, one often drops the subscript and writes, with abuse of notation, T^{-1} also for a pseudo-inverse.

The effect of a linear mapping on a subspace is determined by the following proposition.

Proposition 2.5. *Let* $U \subset V$ *be a linear subspace of the linear space* V, *and let* $T : V \to W$ *be a linear map from* V *to the linear space* W. *Then*

i) *the image* $T(U) \subset W$ *is a linear subspace;*

ii) *the pre-image of a linear subspace* $K \subset W$ *under* T *is a linear subspace of* V.

Further insight into linear maps can be obtained if we endow the space of all such maps with additional structure.

Theorem 2.1. *The set of all linear maps from a space* V *into a space* W *is the **linear space** $L(V, W)$ **of linear maps from** V **to** W with addition*

$$(A + B)(x) = A\,x + B\,x$$

for $A, B \in L(V, W)$ *and* $x \in V$, *and scalar multiplication*

$$(a\,B)(x) = a\,B(x).$$

Next to addition and scalar multiplication, another useful operation for linear maps is the following.

Definition 2.16. *Let* $A : U \to V$ *and* $B : V \to W$ *be linear maps. The **composition** $B \circ A : U \to W$ is*

$$(B \circ A)(x) = B(A(x)).$$

It is important to note that composition of linear maps is **not commutative**, and the situations when it does commute are highly non-generic. The following definition introduces an important characteristics of linear maps.

Definition 2.17. *Let* $T : V \to W$ *be a linear map. The **nullspace** or **kernel** of* T *is*

$$\ker(T) = \{x \in V \mid Tx = 0\}.$$

*The **range** of* T *is the image of* V *under* T, *that is*

$$\operatorname{ran}(T) = \{y \in W \mid Tx = y, x \in V\}.$$

Intuitively, the kernel is the subset of V which has no representation under the mapping T. This is illustrated by the following example.

Example 2.8. Let $U = V \oplus W$ be the direct sum of two linear spaces V and W, and let a linear operator $P : U \to U$ be defined as

$$P : (v, w) \mapsto (v, 0).$$

The kernel $\ker(P)$ of P is given by

$$\ker(P) = (0, w) \cong W$$

and the range $\operatorname{ran}(P)$ is

$$\operatorname{ran}(P) = (v, 0) \cong V.$$

An operator with the above properties, which splits a space into orthogonal complements, is a **projection operator**, and we will encounter them again for example in the context of bases for vector spaces.

Important properties of the nullspace and range of an operator are summarized in the following proposition.

Proposition 2.6. *Let $A : U \to V$ and $B : V \to W$ be linear maps. Then*

i) the kernel $\ker(A) \subset U$ and the range $\operatorname{ran}(A) \subset V$ of A are linear subspaces;

ii) A is invertible if and only if $\ker(A) = \{0\}$ and $\operatorname{ran}(A) = V$;

iii) if the composition $B \circ A$ is invertible then $\ker(A) = \{0\}$ and $\operatorname{ran}(B) = W$;

iv) if A and B are invertible so is their composition, and the inverse is

$$(B \circ A)^{-1} = A^{-1} \circ B^{-1}.$$

It should be noted how the order of the maps is interchanged for the inverse of a composition. A linear map that will play an outstanding role in subsequent discussions is the following.

Definition 2.18. *Let V be a linear space. A **linear functional** Y is a linear map $Y : V \to \mathbb{R}$ from V into the real numbers $\mathbb{R}$. The space $L(V, \mathbb{R})$ of all linear functionals is the **algebraic dual space** V^+ of V..*

Remark 2.2. V^+ is referred to as algebraic dual to distinguish it from the continuous dual V^* that will be introduced later when we have a norm at our disposal and continuity has been introduced.

Many important properties of the dual space follow immediately from the fact that it is a vector space of linear maps. The following proposition characterizes the dual of a dual space.

Proposition 2.7. *Let V be a linear space with dual V^+. The **double dual** $V^{++} = (V^+)^+$ of V is naturally homomorphic to V with the homomorphism $f : V \to V^{++}$ given by*

$$f(x)(Y) = Y(x)$$

for $x \in V$ and $Y \in V^*$.

For the algebraic dual, V and V^{++} are isomorphic if and only if V is finite dimensional, but, as we will see in the following, the situation is rather different for the continuous dual. Using the notion of a dual space, we can introduce the transpose or adjoint of a linear map.

Definition 2.19. *Let $T : V \to W$ be a linear map between linear spaces* V *and* W, *and let* V^+ *and* W^+ *be the respective dual spaces. The **transpose** or **adjoint** T^* of T is the linear map $T^* : W^*$ given by*

$$(T^*(Y))(x) = Y(T(x))$$

for $x \in V$ and $Y \in W^+$.

It should be noted that the transpose T^* is a map in the inverse direction of T. We will encounter the map again in Chapter 2.3 when we introduce the pullback.

2.2.1.3 Banach Space

The linear spaces we considered so far were rather "plain". In the following, we will furnish them with additional useful structure, which will subsequently lead to the concept of a Banach space.

Definition 2.20. *Let $\mathbf{E}$ be a linear space. A **norm** $\| \cdot \|$ **on** $\mathbf{E}$ *is a function* $\| \cdot \| : \mathbf{E} \to \mathbb{R}$ *into the real numbers which for all $x, y \in \mathbf{E}$ satisfies*

 i) *positivity, $\|x\| > 0$, and $\|x\| = 0$ for $x = 0$;*

 ii) *triangle inequality, $\|x + y\| < \|x\| + \|y\|$;*

 iii) *homogeneity, $\|a\,x\| = |a|\,\|x\|$ for all scalars $a \in \mathbb{R}$.*

*A linear space $\mathbf{E}$ together with a norm $\| \cdot \|$ is a **normed space** $(\mathbf{E}, \| \cdot \|)$.*

When the norm is clear from the context, we will often only write $\mathbf{E}$ to refer to the tuple $(\mathbf{E}, \| \cdot \|)$. Intuitively, a norm provides a way to "measure things" in a vector space, and the above requirements for a norm ensure that it behaves like the usual norm in Euclidean space.

Example 2.9. The Lebesgue space $L_1([0, 1])$ on the unit interval $[0, 1] \subset \mathbb{R}$ is

$$L_1([0, 1]) = \{f : [0, 1] \to \mathbb{R} \mid \|f\|_1 < \infty\} \tag{2.1}$$

where the L_1 norm $\| \cdot \|_1$ is defined by

$$\|f\|_1 = \int_0^1 |f(x)| \ dx. \tag{2.2}$$

L_1 is hence a normed space, whose vector space structure is those from Example 2.5. We will consider Lebesgue spaces in more detail in later examples since these have additional structure.

Remark 2.3. A space similar to a normed space $(\mathbf{E}, \|\cdot\|)$ is a metric space (S, d). However, a metric space need not to be linear and a metric is a bilinear operation $d : S \times S \to \mathbb{R}$. When S is a linear space and the metric d is translation invariant and homogeneous, then (S, d) is a normed space. Conversely, every normed space $(\mathbf{E}, \|\cdot\|)$ is a metric space with metric $d(x, y) = \|x - y\|$ for $x, y \in \mathbf{E}$.

A particularly tractable and useful class of normed spaces are separable ones. In such spaces one can obtain constructive expressions for many concepts that otherwise remain abstract definitions.

Definition 2.21. *Let $\mathbf{E}$ be a normed space and $\mathbf{F} \subset \mathbf{E}$ be a subset of $\mathbf{E}$. Then $\mathbf{F}$ is **dense** when the closure of $\mathbf{F}$ is $\mathbf{E}$.[6] A normed space $\mathbf{E}$ that contains a countable set of elements that is dense is called **separable**.*

Unless mentioned otherwise, in the following we will always assume that our spaces are separable. The norm that is available on normed spaces also provides additional structures for maps between such spaces. A crucial notion in this context is that of continuity.

Definition 2.22. *Let $A : \mathbf{E} \to \mathbf{F}$ be a linear map between normed space $(\mathbf{E}, \|\cdot\|_{\mathbf{E}})$ and $(\mathbf{F}, \|\cdot\|_{\mathbf{F}})$. Then T is **continuous** or **bounded** if there exists a positive, real constant $M > 0$ such that for all $x \in \mathbf{E}$ it holds*

$$\|A\,x\|_{\mathbf{F}} \le M\,\|x\|_{\mathbf{E}}.$$

The space of all continuous linear maps from $\mathbf{E}$ to $\mathbf{F}$ is denoted by $L^c(\mathbf{E}, \mathbf{F})$.

From the above definition it is apparent why continuous linear maps are also known as bounded maps.

Remark 2.4. The concept of continuity does not require a norm or a metric, but only a topology on a set is required. However, for our purposes the above definition suffices.

The following proposition will be of much importance in the following in the context of numerical computations.

Proposition 2.8. *Let $A : \mathbf{E} \to \mathbf{F}$ be a linear map between normed spaces, and let $\mathbf{E}$ be finite dimensional. Then A is continuous.*

A norm also allows us to introduce the concept of completeness, which in many contexts is required for a normed space to be "well behaved". Before we define the concept, however, let us briefly recall some elementary notions of convergence.

---

[6]We have to break at this point with our promise to not include topology in the present discussion. The closure of $A \subset S$, where S is a topological space, is the intersection of all closed sets containing A (Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*, Def. 1.1.7).

Remark 2.5. Let **E** be a normed vector space, and $\{x_i\}_{i=1}^{\infty}$ be a sequence in **E**. Then $\{x_i\}_{i=1}^{\infty}$ converges to $x \in \mathbf{E}$ when

$$\|x - x_i\| \xrightarrow[i \to \infty]{} 0, \tag{2.3}$$

and it is a Cauchy sequence if for every $\epsilon > 0$ there exists an $n \in \mathbb{N}$ such that

$$\|x_k - x_l\| \leq \epsilon \, , \ k, l \geq n. \tag{2.4}$$

With these notions again in mind, we can define when a normed space is complete.

Definition 2.23. *A normed space* $(\mathbf{E}, \| \cdot \|)$ *is* **complete** *if every Cauchy sequence in* **E** *has a limit in* **E**.

Completeness hence guarantees that limits are well defined within the space, and taking a limit cannot take us out of the space.

Example 2.10. The open interval $(0, 1) \subset \mathbb{R}$ with the usual absolute value as norm is not complete. The closed interval $[0, 1] \subset \mathbb{R}$ with the same norm is complete.

Remark 2.6. Given a metric or a normed space, it is always possible to complete the space such that the original space is dense in the completion. We will not consider the rather technical aspects of completion, but it is useful to keep the fact in mind.

Normed spaces that are complete are of great importance in many settings and deserve their own name.

Definition 2.24. *A* **Banach space** *is a complete normed space.*

We will now turn again to linear spaces and study the additional structure that is induced by a norm. For convenience, we will in the following always assume that our spaces are Banach, even when completeness is not strictly required for the concepts that will be introduced. An important example of Banach spaces is discussed in the following.

Example 2.11. The rational numbers $\mathbb{Q}$ under the absolute value as norm are not complete. The completion of $\mathbb{Q}$ are the real numbers $\mathbb{R}$.

Remark 2.7. A linear map $A : \mathbf{E} \to \mathbf{F}$ between Banach spaces is often referred to as an **operator**, in particular in the case when it is an endomorphism, that is when $A : \mathbf{E} \to \mathbf{E}$.

The natural linear mapping in a Banach space is the following.

Definition 2.25. *Let* $(\mathbf{E}, \| \cdot \|_{\mathbf{E}}$ *and* $(\mathbf{F}, \| \cdot \|_{\mathbf{F}})$ *be Banach spaces. An* **isometry** f *is a map* $f : \mathbf{E} \to \mathbf{F}$ *which for all* $x \in \mathbf{E}$ *satisfies*

$$\|f(x)\|_{\mathbf{F}} = \|x\|_{\mathbf{E}}.$$

An isometry is hence a norm preserving mapping, and obviously and important special case is that of an endomorphism when f maps $\mathbf{E}$ into itself. Note that in general an isometry need not to be linear. The special case when this is satisfied is given by the following theorem, a result apparently due to Mazur and Ulam.[7]

Theorem 2.2. *Let $(\mathbf{E}, \|\cdot\|_{\mathbf{E}})$ and $(\mathbf{F}, \|\cdot\|_{\mathbf{F}})$ be Banach spaces and $f : \mathbf{E} \to \mathbf{F}$ an isometry. When f maps the origin of $\mathbf{E}$ to the origin of $\mathbf{F}$, then it is a linear map.*

With a norm for elements in a linear space, we can also introduce a norm for operators acting on the space.

Definition 2.26. *Let $(\mathbf{E}, \|\cdot\|)$ and $(\mathbf{F}, \|\cdot\|)$ be Banach spaces, and $A : \mathbf{E} \to \mathbf{F}$ be a bounded linear operator. Then the **operator norm** of A is*

$$\|A\| = \sup \left\{ \frac{\|Ax\|}{\|x\|} \;\middle|\; x \in \mathbf{E}, x \neq 0 \right\}.$$

Note that different operator norms exist, and one has to be careful to distinguish them. Similar to the fact that the space of linear operators on a vector space forms itself a vector space, we have the following result.

Proposition 2.9. *Let $(\mathbf{E}, \|\cdot\|)$ and $(\mathbf{F}, \|\cdot\|)$ be Banach spaces. Then the space $L(\mathbf{E}, \mathbf{F})$ of linear maps from $\mathbf{E}$ to $\mathbf{F}$ is a normed space $(L(\mathbf{E}, \mathbf{F}), \|\cdot\|)$ with the operator norm $\|\cdot\|$, and when $\mathbf{F}$ is a Banach space, then so is $(L(\mathbf{E}, \mathbf{F}), \|\cdot\|)$.*

Recall from the previous paragraph that the algebraic dual space V^+ of a linear space V, the space of linear functionals on V, was given by $V^+ = L(V, \mathbb{R})$. The subspace of V^+ formed by continuous linear functionals in the sense of Def. 2.22 is the continuous dual.

Definition 2.27. *Let $\mathbf{E}$ be a Banach space. The **continuous dual space** $\mathbf{E}^*$ of $\mathbf{E}$ is the space of continuous linear functionals $A \in L^c(\mathbf{E}, \mathbb{R})$.*

It is important to note that $\mathbf{E}^* \subset \mathbf{E}^+$, and in fact it is a linear subspace of $\mathbf{E}^+$. Unless mentioned otherwise, we will in the following always employ the continuous dual space. In contrast to the algebraic dual, the continuous dual is isomorphic to the original space V not only when V is finite dimensional. The important situation when this is the case deserves a name on its own.

Definition 2.28. *A Banach space $\mathbf{E}$ is **reflexive** when $\mathbf{E} \cong \mathbf{E}^{**}$, and the continuous double dual space $\mathbf{E}^{**}$ is isomorphic to $\mathbf{E}$.*

In the following, we will consider examples of Banach spaces which will also play a role in the remainder of the thesis.

---

[7]Mazur and Ulam, "Sur les transformations isométriques d'espaces vectoriels normés".

Example 2.12. In Example 2.9 we introduced the Lebesgue space $L_1([0,1])$. For $1 < p < \infty$, the definition is naturally extended to the family of **Lebesgue spaces** L_p defined by

$$L_p = \{f : [0,1] \to \mathbb{R} \mid \|f\|_p < \infty\} \tag{2.5}$$

where the p-norm $\|\cdot\|_p$ is

$$\|f\|_p = \left(\int_0^1 |f(x)|^p \, dx\right)^{1/p}. \tag{2.6}$$

The so defined L_p spaces are Banach spaces,[8] and they are of considerable importance in applications, for example in physics and signal processing. One of the reasons for the theoretical and practical importance of Lebesgue spaces is that the continuous dual spaces L_p^* can easily be identified. With

$$\frac{1}{q} + \frac{1}{p} = 1 \tag{2.7}$$

there is an isomorphism $\varphi : L_p^* \to L_q$ given by

$$F(g) = \int \varphi(F)(x) \, g(x) \, dx \tag{2.8}$$

for $F \in L_p^*$ and $g \in L_q$. Hence, $L_p^* \cong L_q$ and the continuous dual space L_p^* is isomorphic to L_q. Note that it follows from Eq. 2.7 that the space L_2 is isomorphic to itself, that is, it is reflexive. This points to some additional structure in the space, and we will hence encounter it again in the next section. It has to be noted that Lebesgue spaces can be defined over arbitrary measure spaces (X, Σ, μ) where integration is well defined.

Example 2.13. Sobolev spaces play an important role in the theory of partial differential equations. They are modelled on the Lebesgue spaces L_p introduced in the foregoing example but instead of requiring that only the function values are "well behaved" one also imposes the constraint on the derivatives. For a suitable set X, the **Sobolev space** $W^{k,p}(X)$ is thus defined as

$$W^{k,p}(X) = \{f \in L_p(X) \mid (D^\alpha f) \in L_p(X), \forall |\alpha| \le k\} \tag{2.9}$$

and it requires that all mixed derivatives D^α whose total order $|\alpha|$ is at most k lie in the Lebesgue space $L_p(X)$, where α is a multi-index. When suitably completed, the Sobolev space $W^{k,p}$ forms a Banach space whose norm is given by

$$\|f\|_{k,p} = \left(\sum_{|\alpha| \le k} \left(\|D^\alpha f\|_p\right)^p\right)^{1/p} \tag{2.10}$$

---

[8]Depending on the notion of integration used, one might have to complete the space by taking a quotient with respect to the kernel of $\|\cdot\|_p$. For Lebesgue integration, the completeness of L_p is known as Riesz-Fischer theorem.

where the summation is over all mixed derivatives of at most order k. When $X = \mathbb{R}^n$, then the Sobolev spaces $W^{k,p}(\mathbb{R}^n)$ can also be defined using the Fourier transform F as

$$W^{k,p}(\mathbb{R}^n) = \left\{ f \in L_p(\mathbb{R}^n) \mid F^{-1}\left((1+\xi)^{k/2} \hat{f} \right) \in L_p(\mathbb{R}^n) \right\}, \qquad (2.11)$$

where $\hat{f}$ denotes the Fourier transform of the function f, that is $\hat{f} = F(f)$, and the behaviour of the derivatives is controlled by the weight $(1+\xi)^{k/2}$ which enforces a suitable decay of the Fourier coefficients as the frequency ξ goes to infinity. The above definition can be extended from $\mathbb{R}^n$ to other domains where a suitable generalization of the Fourier transform is defined, for example on the sphere,[9] and it also allows to introduce Sobolev spaces where k is not an integer.

Example 2.14. In applications where discontinuities arise a well behaved class of functions are whose for which the variation

$$V_a^b(f) = \int_a^b |f'(x)| \, dx \qquad (2.12)$$

is finite when a suitable notion of differentiation is employed. The **space of functions with bounded variation** $\mathrm{BV}([a,b])$ is hence defined as

$$\mathrm{BV}([a,b]) = \left\{ f \in L_1([a,b]) \mid V_a^b(f) < \infty \right\} \qquad (2.13)$$

and it is a linear subspace of $L_1([a,b])$. Moreover, with the norm

$$\|f\|_{\mathrm{BV}} = \|f\|_1 + V_a^b(f) \qquad (2.14)$$

the space $\mathrm{BV}([a,b])$ is a Banach space. In higher dimensions and over more complex domains, the space of functions of bounded variation can be defined using distributional derivatives. It should be noted that the space of functions of bounded variation is not separable, contrary to the promise we in the foregoing to only consider such spaces.

The dual space $\mathbf{E}^*$ of a Banach space provides a natural pairing $\mathbf{E} \times \mathbf{E}^* \to \mathbb{R}$ between elements in the spaces by $Y(x) \in \mathbb{R}$ for $x \in \mathbf{E}$ and $Y \in \mathbf{E}^*$. Next, we will introduce a slightly more general notion of a pairing between two arbitrary Banach spaces, which will also be useful later on in the context of geometric mechanics.

Definition 2.29. *Let* $\mathbf{E}$ *and* $\mathbf{F}$ *be Banach spaces. A* ***pairing*** $\langle\,,\rangle$ *between* $\mathbf{E}$ *and* $\mathbf{F}$ *is a continuous, bilinear map*

$$\langle\,,\rangle : \mathbf{E} \times \mathbf{F} \to \mathbb{R}.$$

---

[9]For Sobolev spaces over the sphere see for example (Freeden, Gervens, and Schreiner, *Constructive Approximation on the Sphere (With Applications to Geomathematics)*; Hesse, "Complexity of numerical integration over spherical caps in a Sobolev space setting").

*The pairing is **E-weakly non-degenerate** when $\langle x, y \rangle = 0$ for all $y \in \mathbf{F}$ implies that $x = 0$, and it is **F-weakly non-degenerate** when $\langle x, y \rangle = 0$ for all $x \in \mathbf{E}$ implies $y = 0$. A pairing $\langle\,,\,\rangle : \mathbf{E} \times \mathbf{F} \to \mathbb{R}$ is weakly non-degenerate when it is both **E**-weakly non-degenerate and **F**-weakly non-degenerate. The Banach spaces **E** and **F** are then in **duality**.*

We present an alternative characterization of weakly non-degenerate pairings in the following proposition.

Proposition 2.10. *Let $\mathbf{E}$ and $\mathbf{F}$ be Banach spaces. Then an $\mathbf{E}$-weakly non-degenerate pairing $\langle\,,\,\rangle : \mathbf{E} \times \mathbf{F} \to \mathbb{R}$ between $\mathbf{E}$ and $\mathbf{F}$ defines a map $\pi_{\mathbf{E}} : \mathbf{E} \to \mathbf{F}^*$ given by*

$$\langle x, y \rangle = \pi_{\mathbf{E}}(x)(y)$$

for $x \in \mathbf{E}$ and $y \in \mathbf{F}$. Analogously, an $\mathbf{F}$-weakly non-degenerate pairing $\langle\,,\,\rangle : \mathbf{E} \times \mathbf{F} \to \mathbb{R}$ between $\mathbf{E}$ and $\mathbf{F}$ defines a map $\pi_{\mathbf{F}} : \mathbf{F} \to \mathbf{E}^*$ given by

$$\langle x, y \rangle = \pi_{\mathbf{F}}(y)(x).$$

The maps $\pi_{\mathbf{E}}$ and $\pi_{\mathbf{F}}$ are homomorphisms.

With the above proposition it is natural to introduce the notion of a strongly non-degenerate pairing which extends the homomorphisms $\pi_{\mathbf{E}}$ and $\pi_{\mathbf{F}}$ to isomorphisms.

Definition 2.30. *Let $\mathbf{E}$ and $\mathbf{F}$ be Banach spaces. A pairing $\langle\,,\,\rangle : \mathbf{E} \times \mathbf{F} \to \mathbb{R}$ is **strongly non-degenerate** when the induced maps $\pi_{\mathbf{E}} : \mathbf{E} \to \mathbf{F}^*$ and $\pi_{\mathbf{F}} : \mathbf{F} \to \mathbf{E}^*$ are isomorphisms.*

Example 2.15. Let $\mathbf{E}$ be a Banach space and $\mathbf{E}^*$ its continuous dual. Then $\mathbf{E}$ and $\mathbf{E}^*$ are in duality and

$$\langle\,,\,\rangle : \mathbf{E} \times \mathbf{E}^* \to \mathbb{R}$$

is a weakly non-degenerate pairing. When $\mathbf{E}$ is reflexive, for example $\mathbf{E}$ is finite dimensional, then the pairing is strongly non-degenerate.

Shortly, we will introduce bases for Banach spaces. We will begin by considering the convergence of infinite linear combinations $\sum_{i=1}^{\infty} x_i$ in this setting.

Definition 2.31. *Let $\mathbf{E}$ be a Banach space, and let $\{x_i\}_{i=1}^{\infty}$ be a sequence of elements of $\mathbf{E}$. The infinite linear combination $\sum_{i=1}^{\infty} x_i$ defined by the sequence $\{x_i\}_{i=1}^{\infty}$ **converges to** $y \in \mathbf{E}$ when*

$$\left\| y - \sum_{i=1}^{n} x_i \right\| \xrightarrow[n \to \infty]{} 0$$

and in this case one writes

$$y = \sum_{i=1}^{n} x_i.$$

The sequence $\{x_i\}_{i=1}^{\infty}$ is **unconditionally convergent** when the infinite series $\sum_{i=1}^{n} x_{\sigma(i)}$ converges for all permutations σ of the natural numbers,[10] and it is **absolutely convergent** when

$$\sum_{i=1}^{\infty} \|x_i\| < \infty.$$

In the following, it will be important to keep in mind that the equality in a series expansion $y = \sum_{i=1}^{n} x_i$ stands for convergence in the norm as the number of terms in the series goes to infinity.

Before, we considered linear combinations of elements in a vector space already in Proposition 2.3 where we employed them to characterize the span of the subset of a vector space. An immediate consequence of the definition of completeness of a sequence of elements in a linear space in Def. 2.11 together with the definition of convergence in Def. 2.31 is the following.

Corollary 2.1. *Let $\{x_i\}_{i=1}^{\infty}$ be a complete sequence for the Banach space $\mathbf{E}$. Then any element y in $\mathbf{E}$ can be approximated arbitrarily well by $\{x_i\}_{i=1}^{\infty}$, that is for every $\epsilon > 0$ there exists and $n \in \mathcal{N}$ and a finite index set I of cardinality n such that for real coefficients $a_i \in \mathbb{R}$ it holds*

$$\left\| y - \sum_{i \in I} a_i\, x_i \right\| \le \epsilon.$$

The above corollary tell us *that* we can approximate an element in a Banach space with some sequence, but it is non-constructive in that it does not specify *how* one can determine a suitable sequence nor the coefficients a_i. The crucial concept of a basis will overcome these problems and it is introduced next.

Bases for Banach Spaces   Bases enable to represent an arbitrary element in a Banach space as a linear combination of representative elements with expansion coefficients that are well defined. They are hence central to the study of spaces that admit such a representation.

Definition 2.32. *Let $\mathbf{E}$ be a Banach space. A sequence $\{e_i\}_{i=1}^{k}$ of elements $e_i \in \mathbf{E}$, with k possibly being infinity, is a **Schauder basis of** $\mathbf{E}$ if for every $y \in \mathbf{E}$ there exists a unique sequence of scalar coefficients $\{a_i(y)\}_{i=1}^{k}$ such that the **basis expansion***

$$y = \sum_{i=1}^{k} a_i(y)\, e_i$$

*holds. If $\{e_{\sigma(i)}\}_{i=1}^{k}$ provides a basis for every permutation σ of the natural numbers then $\{e_i\}_{i=1}^{k}$ is an **unconditional Schauder basis**.*

---

[10]For a discussion of permutations the reader is referred to Remark 2.62.

A Schauder basis hence allows to *uniquely* represent each element in $\mathbf{E}$ as a linear combination of a set of elements $e_i \in \mathbf{E}$. As usual, the equality in the above definition should be understood in the sense of a convergent linear combination in Def. 2.31. Unless confusion can arise, in the following we will often write 'basis' when we refer to a Schauder basis.

Remark 2.8. In the foregoing, we posited that we are working in separable Banach spaces. One of the reasons for this was that every Banach space that has a basis, and hence in a certain sense every linear space that is useful to us, is separable. However, the converse is not true and there are separable Banach spaces which do not admit a basis, apparently a rather recent result by Enflo.[11]

The basis functions $e_i \in \mathbf{E}$ of a Schauder basis have the following important property, which follows from the uniqueness of the coefficients $a_i(y) \in \mathbb{R}$, cf. Def. 2.11.

Proposition 2.11. *Let $\mathbf{E}$ be a Banach space and $\{e_i\}_{i=1}^k$ be a Schauder basis for $\mathbf{E}$. Then the $e_i \in \mathbf{E}$ are linearly independent.*

With a basis, an alternative and more practical characterization of the dimension of a Banach space is the following.

Corollary 2.2. *Let $\mathbf{E}$ be a Banach space and $\{e_i\}_{i=1}^k$ a basis for $\mathbf{E}$. When the cardinality $\mathrm{card}\,(\{e_i\})$ of the basis is finite, that is $k < \infty$, then*

$$\dim\,(\mathbf{E}) = \mathrm{card}\,(\{e_i\})$$

and the dimensionality of $\mathbf{E}$ is given by the cardinality of $\{e_i\}_{i=1}^k$, with the isomorphism being provided by the coordinate functionals and the basis functions. Otherwise, $\mathbf{E}$ is infinite dimensional.

An important characterization of Schauder bases can be obtained with the following notion, which a priori is defined for arbitrary sequences in a Banach space.

Definition 2.33. *Let $\mathbf{E}$ be a Banach space and $\{x_i\}_{i=1}^k$ be a sequence of elements $x_i \in \mathbf{E}$. The **basis constant of** $\{x_i\}_{i=1}^k$ is*

$$K = \sup \left\{ \left\| \sum_{i=1}^m a_i\, x_i \right\| \; : \; \left\| \sum_{i=1}^n a_i\, x_i \right\| = 1\,,\, m \leq n \right\}$$

with the supremum over all finite, real-valued sequences $\{a_i\}_{i=1}^k$.

Intuitively, the basis constant describes how much a linear combination of the sequence $\{e_i\}_{i=1}^k$ can "blow up". This is made precise in the following proposition.

---

[11]Enflo, "A counterexample to the approximation problem in Banach spaces".

Proposition 2.12. *Let* **E** *be a Banach space and* $\{e_i\}_{i=1}^k$ *a sequence of elements* $e_i \in \mathbf{E}$. *Then* $\{e_i\}_{i=1}^k$ *is a basis if and only if the basis constant* K *is finite, that is* $K < \infty$.

In the definition of a Schauder basis in Def. 2.32, the existence of the scalar coefficients $a_i(y)$ was posited, but there we did not provide an explanation of how these can be obtained for a given element $y \in \mathbf{E}$. We will now characterize the $a_i(y)$. A crucial observation is the following corollary to Proposition 2.12.

Proposition 2.13. *Let* **E** *be a Banach space and* $\{e_i\}_{i=1}^k$ *a Schauder basis for* **E**. *Then the* ***coefficient functionals*** a_i *defined by*

$$y = \sum_{i=1}^k a_i(y)\, e_i$$

are continuous functionals $a_i \in \mathbf{E}^* = L^c(\mathbf{E}, \mathbb{R})$. *Moreover, if there exists a constant* $C > 0$ *such that all basis functions* e_i *satisfy* $\|e_i\| < C$, *then for all continuous linear functionals* a_i *the norm* $\|a_i\|$ *is uniformly bounded.*

The above proposition associates with a basis $\{e_i\}_{i=1}^k$ for **E** a sequence $\{a_i\}_{i=1}^k$ in the dual space $\mathbf{E}^*$, and the dual elements $a_i \in \mathbf{E}^*$ determine the scalar coefficients needed for a basis expansion as $a_i(y)$. Before we study the duality between $\{e_i\}_{i=1}^k$ and $\{a_i\}_{i=1}^k$ in more detail, two remarks are in order.

Remark 2.9. In Proposition 2.13, we identified a map $a_i : \mathbf{E} \to \mathbb{R}$ with its image $a_i(y) \in \mathbb{R}$. Which of the two facets we refer to will always be clear from the context, and we will employ similar identifications more often in the following.

Remark 2.10. A coefficient functional a_i together with the associated basis function e_i forms the projection operator P_i onto the subspace spanned by the ith basis functions, that is

$$P_i(y) = a_i(y)\, e_i.$$

Hence, a collection of coefficient functionals projects onto the linear span of the associated basis functions, see also Example 2.8 and Proposition 2.3.

The relationship between a basis $\{e_i\}_{i=1}^k$ in **E** and the sequence $\{a_i\}_{i=1}^k$ in $\mathbf{E}^*$ will be formalized using the notion of a biorthogonal system, for which we employ the pairing between two Banach spaces from Def. 2.29.

Definition 2.34. *Let* **E** *be a Banach space and* $\mathbf{E}^*$ *its (continuous) dual space, and let* $I \subseteq \mathbb{N}$ *be a non-empty index set. A* ***biorthogonal system*** *is a sequence of tuples* $\{(e_i, \tilde{e}_i)\}_{i=1}^k$ *in* $\mathbf{E} \times \mathbf{E}^*$ *such that*

$$\tilde{e}_j(e_i) = \langle e_j, \tilde{e}_i \rangle = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

and the sequences $\{e_i\}_{i=1}^k$ *and* $\{\tilde{e}_j\}_{j=1}^k$ *are then said to be* ***biorthogonal***.

The δ_{ij} in the above definition is known as Kronecker delta, and we will use it frequently in the following.

We are now prepared to characterize the coefficients functionals and their relationship to the basis for which they are defined.

Theorem 2.3. *Let* $\mathbf{E}$ *be a Banach space, and let* $\{e_i\}_{i=1}^k$ *be a basis for* $\mathbf{E}$ *with coefficient functionals* $\{\tilde{e}_j\}_{j=1}^k$ *in* $\mathbf{E}^*$, *with* k *possibly being infinity. Then* $\{e_i\}_{i=1}^k$ *and* $\{\tilde{e}_i\}_{i=1}^k$ *form a biorthogonal system and* $\{\tilde{e}_j\}_{j=1}^k$ *forms a **dual basis** for its closed span in* $\mathbf{E}^*$. *Moreover, when* $\mathbf{E}^*$ *is reflexive then* $\{\tilde{e}_j\}_{j=1}^k$ *is a dual basis for* $\mathbf{E}^*$.

The above theorem hence asserts that when $\mathbf{E}$ is reflexive, which will usually be the case in the following, then ***the basis*** $\{e_i\}_{i=1}^k$ ***determines a natural dual basis*** $\{\tilde{e}_j\}_{j=1}^k$ ***for*** $\mathbf{E}^*$ ***by the biorthogonal condition***

$$\langle e_j, \tilde{e}_i \rangle = \delta_{ij},$$

and, moreover, the dual basis provides the coefficients needed for the basis expansion of an arbitrary element $y \in \mathbf{E}$ ***with respect to*** $\{e_i\}_{i=1}^k$. Although we will usually assume slightly more structure than is available in a Banach space and a slightly less general notion of a basis will be employed, the construction just stated will be our recipe to effectively work with arbitrary functions.

2.2.2 Hilbert Spaces

The norm of a Banach space, or more generally a normed space, provides a means to measure distances and length. However, it does not enable to determine the relative relationship of two elements in such a space. In the following, we will introduce the inner product that enables to characterize "angles" in vector spaces, and Hilbert spaces will be obtained as the completion of spaces with such an inner product.

2.2.2.1 Inner Product Spaces and Hilbert Spaces

We will begin by introducing inner products, the generalization of the scalar product from Euclidean space to arbitrary vector spaces.

Definition 2.35. *Let* V *be a linear space. An **inner product** $\langle \, , \rangle : \mathrm{V} \times \mathrm{V} \to \mathbb{R}$ is a function into the reals which for all* $x, y \in \mathrm{V}$ *satisfies*

   i) bilinearity, for fixed x, *the inner product* $\langle x, y \rangle$ *is linear in* y, *and for fixed* y, *it is linear in* x;

   ii) symmetry, $\langle x, y \rangle = \langle y, x \rangle$;

   iii) positivity, $\langle x, x \rangle > 0$ *for* $x \neq 0$.

A linear space V *together with an inner product* $\langle \, , \rangle$ *is an **inner product space** $(V, \langle \, , \rangle)$.*

The inner product is the generalization of the scalar product in Euclidean space $\mathbb{R}^n$, and its properties are modelled on those of the scalar product.

Remark 2.11. An inner product provides a strongly non-degenerate pairing between a vector space V and itself.

The following proposition shows that every inner product space is also a normed space.

Proposition 2.14. *Let* $(V, \langle , \rangle)$ *be an inner product space. Then* V *is a normed space with induced norm*

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Unless stated otherwise, the norm we use in the context of inner product spaces will always be the natural or induced norm from the above proposition. Analogous to Euclidean space, orthogonality between two vectors is defined as a vanishing inner product.

Definition 2.36. *Let* $\mathcal{H}$ *be a Hilbert space. Two elements* $x, y \in \mathcal{H}$ *are **orthogonal** when*

$$\langle x, y \rangle = 0.$$

A result of considerable importance for inner product spaces, which for example will enable to bound errors in the sequel, is the following.

Theorem 2.4 (Cauchy-Schwarz inequality)**.** *Let* $(V, \langle , \rangle)$ *be an inner product space. Then for all* $x, y \in V$ *is holds*

$$|\langle x, y \rangle| \leq \|x\| \, \|y\|.$$

The Cauchy-Schwarz inequality provides a generalization of the equality

$$\vec{x} \cdot \vec{y} = \cos\theta \, \|\vec{x}\| \, \|\vec{y}\|$$

from Euclidean space to arbitrary inner product spaces. Analogous to the construction of a Banach space, the completion of an inner product space yields a space that deserves a name on its own because of its importance in applications.

Definition 2.37. *A **Hilbert space** is a complete inner product space.*

It follows immediately from Proposition 2.14 that every Hilbert space is also a Banach space. In the following, we will consider some classical examples of Hilbert spaces which will also be of importance in the sequel.

Example 2.16. Euclidean space $\mathbb{R}^n$ with the usual scalar product, or dot product, as inner product is a Hilbert space.

Example 2.17. The Lebesgue space $L_2([0,1])$, cf. Example 2.12, is a Hilbert space with inner product

$$\langle f, g \rangle = \int_0^1 f(x)\, g(x)\, dx. \tag{2.15}$$

The definition is naturally extended to $L_2(X)$ for an arbitrary measure space (X, Σ, μ).

Example 2.18. The Sobolev spaces $W^{k,2}(X)$ modelled on the Hilbert spaces $L_2(X)$ are Hilbert spaces, cf. Example 2.13, and the inner product for the spaces is given by

$$\langle f, g \rangle = \sum_{|\alpha| \le k} \langle D^\alpha f, D^\alpha g \rangle \tag{2.16}$$

where the summation is over all multi-indices α that have at most order k, and $\langle\,,\rangle$ is the L_2-inner product of Example 2.17. The spaces area usually denoted as Hilbert-Sobolev spaces $H^s(X) = W^{s,2}(X)$.

Example 2.19. The discrete Lebesgue space ℓ_2 is formed by "finite energy", real-valued infinite sequences

$$\ell_2 = \left\{ a = (a_1, a_2, \ldots) \,\middle|\, a_i \in \mathbb{R}, \sum_{i=1}^\infty a_i < \infty \right\} \tag{2.17}$$

with the inner product for $a, b \in \ell_2$ given by the scalar product

$$\langle a, b \rangle = \sum_{i=1}^\infty a_i\, b_i. \tag{2.18}$$

When the elements of the discrete Lebesgue space are finite sequences, then the spaces will be denoted as ℓ_2^n and we have $\ell_2^n \cong \mathbb{R}^n$. For notational convenience, we will usually write ℓ_2^n, but it should be kept in mind that this is just Euclidean space $\mathbb{R}^n$.

In the foregoing, we already encountered linear functionals in the dual space of a vector space, and we saw their importance for example for a Schauder basis where the dual basis was given by such functionals. In a Hilbert space, linear functionals prove to be even more useful, largely due to the following result.

Theorem 2.5 (Riesz Representation Theorem)**.** *Let $\mathcal{H}$ be a Hilbert space. Then $\mathcal{H}$ and its dual $\mathcal{H}^*$ are isomorphic, and for $x \in \mathcal{H}$ and $y \in \mathcal{H}^*$ the isomorphism is given by*

$$y(x) = \langle x, y \rangle.$$

The importance of the Riesz representation theorem lies in the explicit form it provides for the dual space $\mathcal{H}^*$: it is given by $\mathcal{H}$ itself, with the pairing between $\mathcal{H}$ and $\mathcal{H}^*$ replaced with the inner product. The reader might already anticipate the importance of the result when we consider bases for Hilbert spaces.

Remark 2.12. Often when one *can* identify two structure using an isomorphism, this does not mean one *should* do so, and in fact this often leads to confusion since two distinct concepts appear to be the same, although they still "behave" or "act" differently. In the case of a Hilbert space $\mathcal{H}$, for example, one often has functions which are meant to be paired with other functions, such as dual basis functions. Even with the identification of $\mathcal{H}$ and $\mathcal{H}^*$, it is then useful to keep in mind that these are naturally objects in the dual $\mathcal{H}^*$ and are only for convenience represented as objects in $\mathcal{H}$. Another example is tensor calculus, which we will consider in detail in Chapter 2.3.2, where the distinction between the primal and dual space will in fact be crucial, since objects in the different spaces behave differently under coordinate transformations.

An immediate corollary of the Riesz representation theorem is the following.

Corollary 2.3. *Let $\mathcal{H}$ be a Hilbert space. Then $\mathcal{H}$ is reflexive.*

The above corollary and the fact that the Lebesgue space L_2 is a Hilbert space explains the reflexivity of L_2 which we already observed in Example 2.12.

2.2.2.2 Operators in Hilbert Spaces

With the inner product, a finer understanding of linear operators is possible than it was in the foregoing. However, before we consider properties of operators on Hilbert spaces, let us introduce the natural mapping on such spaces.

Definition 2.38. *Let $\mathcal{H}$ be a Hilbert space. A bounded linear operator $U : \mathcal{H} \to \mathcal{H}$ is **real unitary** when it is surjective and for $x, y \in \mathcal{H}$ satisfies*

$$\langle Ux, Uy \rangle = \langle x, y \rangle .$$

From the above definition and Proposition 2.14 it is apparent that every real unitary operator is also an isometry.

Remark 2.13. In the complex case, that is for a Hilbert space over the field $\mathbb{C}$, an operator with the above properties is a unitary operator, and these are of central importance in quantum mechanics.

Example 2.20. A rotation $R : \mathbb{R}^3 \to \mathbb{R}^3$ can be represented by a real unitary operator that leaves the inner product in Euclidean space invariant. The inverse rotation is given by the transpose so that

$$RR^T = R^T R = I$$

where I is the 3×3 identity matrix.

The above example generalizes as the following proposition shows.

Proposition 2.15. *Let $\mathcal{H}$ be a Hilbert space and $U : \mathcal{H} \to \mathcal{H}$ be an operator. Then U is real unitary if and only if*

$$A^* A = A A^* = I$$

where I is the identity operator on $\mathcal{H}$.

Note that both equalities are needed in the above proposition for the operator to be unitary. A discussion of the space of real unitary operators and its apparently geometric structure will be provided in Chapter 2.3.2.7 and Remark 2.104.

Next, we will employ the inner product to introduce an important class of operators which can be understood as a generalization of real unitary operators, and which are intimately related to them. For this, we will need again the concept of the adjoint or transpose of an operator. For a linear map $T : V \to W$, the adjoint was defined in Def. 2.19 as the linear map $T^* : W^* \to V^*$ such that

$$(T^*(Y))(x) = Y(T(x)),$$

where $x \in V$, $Y \in W^*$. When V, W are Hilbert spaces, then by the Riesz representation theorem this can also be written as

$$\langle Tx, y \rangle = \langle x, T^*y \rangle.$$

With this alternative characterization of the adjoint in a Hilbert space we can introduce the following notion.

Definition 2.39. *Let $\mathcal{H}$ be a Hilbert space and $\mathcal{D} \subset \mathcal{H}$ a dense subspace of $\mathcal{H}$. A bounded, linear operator $A : \mathcal{D} \to \mathcal{D}$ is **self-adjoint** when*

$$\langle Ax, y \rangle = \langle x, Ay \rangle$$

*and it is **anti-self-adjoint** when*

$$\langle Ax, y \rangle = \langle x, -Ay \rangle.$$

Note that it is important that the operator A in the above definition is defined on all of $\mathcal{H}$, or at least on a dense subset thereof, and it is a result by Hellinger and Toeplitz that boundedness of A would not have to be required but that it is a consequence of the other properties. Intuitively, a self-adjoint operator leaves the inner product invariant when going in either direction, from x to Ax and then determining the inner product $\langle Ax, y \rangle$, or from y to Ay and then computing $\langle x, Ay \rangle$.

Another important class of operators, in particular for the applications we have in mind, are Hilbert-Schmidt integral operators.

Definition 2.40. *Let $L_2(X)$ be the Lebesgue-Hilbert space over a measure space (X, Σ, μ). A **Hilbert-Schmidt kernel** is a function $k : X \times X \to \mathbb{R}$ satisfying*

$$\int_X \int_X \|k(x, y)\|^2 dx \, dy < \infty.$$

*The **Hilbert-Schmidt integral operator** $K : L_2(X) \to L_2(X)$ associated with $k(x, y)$ is*

$$g(y) = (Kf)(y) = \int_X f(x) \, k(x, y) \, dx.$$

The condition on the kernel $k : U \times U \to \mathbb{R}$ in Def. 2.40 can be interpreted as requiring that $k(x,y)$ is an element in $L_2(U \times U)$, and this ensures that the operator K and the image $g = (Kf)$ are well behaved.

Remark 2.14. A Hilbert-Schmidt operator is a special type of an integral operator with symmetric kernel defined on $L_2(X)$. The theory can be generalized to more general function space, and then essentially the same results hold.[12]

The following proposition establishes the properties of Hilbert-Schmidt integral operators.

Proposition 2.16. *Let $\mathcal{H}(U)$ be a Hilbert space defined over a compact, connected set $U \subset \mathbb{R}^n$, and let $K : \mathcal{H} \to \mathcal{H}$ be a Hilbert-Schmidt integral operator with Hilbert-Schmidt kernel $k : U \times U \to \mathbb{R}$. Then K is compact, and, moreover if the kernel is symmetric, satisfying $k(x,y) = k(y,x)$, then K is self-adjoint.*

The above theorem states that K is compact, which is a notion from topology, when the image of a bounder subset under K is again a relatively compact subset, and compactness implies for example that K is continuous. As we will see shortly, a precise understanding of this concept is for us not necessary, since we only require it for some subsequent results to apply, and we therefore refer to the literature.[13]

Remark 2.15. A Hilbert-Schmidt operator can be defined without requiring that it is an integral transform, and it is then characterized by the properties in Proposition 2.16.

Next, we will study the spectral properties of Hilbert-Schmidt operators. Before, however, we have to introduce the necessary concepts.

Definition 2.41. *Let $\mathcal{H}$ be a Hilbert space and $A : \mathcal{H} \to \mathcal{H}$ a bounded linear operator, and let I be the identity operator on $\mathcal{H}$. Then the **spectrum** $\sigma(A)$ **of** A is*

$$\sigma(A) = \{\lambda \in \mathbb{C} \mid (\lambda I - A) \text{ is not invertible}\}$$

*and the **resolvent** $r(A)$ **of** A is*

$$r(A) = \{\lambda \in \mathbb{C} \mid (\lambda I - A) \text{ is invertible}\}$$

*and $\sigma(A) \cup r(A) = \mathbb{C}$. The subset of the spectrum $\sigma(A)$ where $(\lambda I - A)$ is not injective is the **point spectrum** $\sigma_p(A)$ **of** A and the elements of $\sigma_p(A)$ are the*

[12]A classical reference on the subject is (Courant and Hilbert, *Methoden der Mathematischen Physik*). Alternative treatments can be found in Smithies (*Integral Equations*) and Wing (*A Primer on Integral Equations of the First Kind*, Chapter 5), and an overview in (Khvedelidze, *Integral Equation with Symmetric Kernel*).

[13]See for example (Stakgold and Holst, *Green's Functions and Boundary Value Problems*, Chapter 5.7).

eigenvalues $\lambda \in \sigma_p(A)$. *The kernel* $\ker(\lambda I - A)$ *is the **eigenspace** associated with* λ, *a vector* $u \in \mathcal{H}$ *in the space satisfying*

$$(\lambda I - A)u = 0$$

*is an **eigenvector** of* A *and one often writes* $Au = \lambda u$. *The dimension of the eigenspace* $\ker(\lambda I - A)$ *is the geometric **multiplicity** m_λ of* $\lambda \in \sigma_p(A)$.

Note that even thought the Hilbert spaces we are considering are defined over the reals, the spectrum is defined in the complex plane.[14] Next to the point spectrum $\sigma_p(A)$, there are other parts of $\sigma(A)$, most notably the continuous spectrum. However, as we will see shortly, for the operators we are interested in, such as Hilbert-Schmidt integral operators, these notions are less of a concern. Returning to such operators, we have the following result.

Theorem 2.6 (Spectral Theorem for Compact Operators)**.** *Let* $\mathcal{H}$ *be a Hilbert space, and let* $A : \mathcal{H} \to \mathcal{H}$ *be a compact and self-adjoint operator on* $\mathcal{H}$. *Then*

  i) *the spectrum consists only of eigenvalues and possibly zero;*

 ii) *all eigenvalues are real;*

iii) *there is at least one and at most a countable number of eigenvalues, with zero the only possible accumulation point;*

 iv) *the eigenspaces are finite dimensional;*

  v) *the eigenvectors in eigenspaces of different eigenvalues are orthogonal.*

The above theorem is also known as Fredholm alternative, since the spectrum consists either of eigenvalues or of zero, and it is part of a theory known as Fredholm theory or Fredholm operator theory, which is concerned with the study of compact operators. It follows immediately from Proposition 2.16 that for a symmetric Hilbert-Schmidt kernel $k(x, y)$ the foregoing theorem applies to the associated Hilbert-Schmidt operator, and the result holds more generally for integral operators with symmetric kernel, cf. Remark 2.14.

Remark 2.16. By its definition, eigenvalues are determined only up to a sign. In the following we will always assume that the eigenvalues λ_i of an operator are ordered such that

$$0 \leq \lambda_1 \leq \ldots \leq \lambda_k$$

and eigenvalues with multiplicity greater than one are repeated.

---

[14]Obviously, this is similar to the fact that the roots of a polynomial are in general complex; a similarity which in fact is not accidental, and which described by the characteristic polynomial.

2.2.2.3 Hilbert Space Bases

In Def. 2.32 we introduced the concept of a Schauder basis for a Banach space **E** which enables to represent an arbitrary element in the space as a linear combination of basis functions with the scalar coefficients, or weights, provided by the dual basis functions in the dual space $\mathbf{E}^*$. Since every Hilbert space is a Banach space, a Schauder basis is also here available. For Hilbert spaces, the additional structure provided by an inner product and the Riesz representation theorem will enable us to develop a refined notion of the concept of a basis. As usual, all Hilbert spaces are assumed to be separable, and when an index sets runs until k, then this is either the dimension of the space if $\mathcal{H}$ is finite dimensional or $k = \infty$ otherwise.

Remark 2.17. For our applications, we are interested in Hilbert spaces that are function spaces. We will hence in the following often refer to an element in a Hilbert space as a function, although the result applies for all Hilbert spaces regardless of the character of the elements.

Before we begin, let us introduce a useful criterion which can be employed to characterize bases, and which will be needed in the sequel.

Definition 2.42. *Let $\mathcal{H}$ be a Hilbert space and $\{\varphi_i\}_{i=0}^k$ be a sequence in $\mathcal{H}$. Then $\{\varphi_i\}_{i=0}^k$ is a **Bessel sequence** when there exists a positive, real constant $B > 0$ such that*

$$\sum_{i=0}^k |\langle f, \varphi_i \rangle|^2 \le B \|f\|^2$$

*for all $f \in \mathcal{H}$. Every constant B satisfying the inequality is then a **Bessel bound** for $\{\varphi_i\}_{i=0}^k$.*

We will see bounds similar to the Bessel bound more often later on, in particular in the context of frames, that is overcomplete bases.

For a reflexive Banach spaces **E** with basis $\{e_i\}_{i=1}^k$, the dual basis $\{\tilde{e}_i\}_{i=1}^k$ was formed by elements in the dual space $\mathbf{E}^*$. Every Hilbert space $\mathcal{H}$ is a reflexive Banach space, and, moreover, using the Riesz representation theorem we can identity the dual functionals $\tilde{e}_i$ with elements in $\mathcal{H}$. This leads to the following result.

Theorem 2.7. *Let $\mathcal{H}$ be a Hilbert space and $\{\varphi_i\}_{i=1}^k$ be a Schauder basis for $\mathcal{H}$. Then the **dual basis** is the unique family $\{\tilde{\varphi}_i\}_{i=1}^k$ of elements $\tilde{\varphi}_i \in \mathcal{H}$ such that for $f \in \mathcal{H}$ one has*

$$f = \sum_{i=1}^k \langle f, \tilde{\varphi}_i \rangle \, \varphi_i.$$

*Moreover, the dual basis $\{\tilde{\varphi}_i\}_{i=1}^k$ also provides a basis for $\mathcal{H}$, and the primal and dual basis form a **biorthogonal basis pair** $(\varphi_i, \tilde{\varphi}_i)_{i=1}^k \equiv (\{\varphi_i\}_{i=1}^k, \{\tilde{\varphi}_i\}_{i=1}^k)$ satisfying*

$$\langle \varphi_i, \tilde{\varphi}_j \rangle = \delta_{ij}.$$
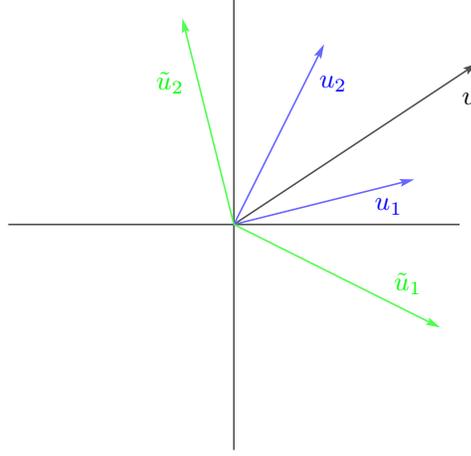
Figure 2.1: Biorthogonal basis for $\mathbb{R}^2$ formed by two vectors u_1 and u_2. Since the vectors are linearly independent, the basis can represent arbitrary vectors $v \in \mathbb{R}^2$.

The above theorem provides a complete characterization of how an arbitrary function $f \in \mathcal{H}$ can be expanded in a basis $\{\varphi_i\}_{i=1}^k$, and how the expansion or basis function coefficients can be obtained using the dual basis $\{\tilde{\varphi}_i\}_{i=1}^k$ by projecting the signal onto each of them using the inner product.

Example 2.21. With respect to the canonical basis (e_1, e_2), a biorthogonal basis for $\mathbb{R}^2$ is given by

$$u_1 = \begin{pmatrix} 1 \\ 1/4 \end{pmatrix} \qquad u_2 = \begin{pmatrix} 1/2 \\ 1 \end{pmatrix}, \tag{2.19}$$

see Fig. 2.1, and the associated dual basis is

$$\tilde{u}_1 = \begin{pmatrix} 8/7 \\ -8/14 \end{pmatrix} \qquad \tilde{u}_2 = \begin{pmatrix} -8/28 \\ 8/7 \end{pmatrix}. \tag{2.20}$$

Indeed, biorthogonality is satisfied since

$$u_1 \cdot \tilde{u}_1 = 1 \cdot 8/7 - 1/4 \cdot 8/14 = 32/28 - 4/28 = 1 \tag{2.21a}$$

$$u_1 \cdot \tilde{u}_2 = -1 \cdot 8/28 + 1/4 \cdot 8/7 = -8/28 + 8/28 = 0 \tag{2.21b}$$

$$u_2 \cdot \tilde{u}_1 = 1/2 \cdot 8/7 - 1 \cdot 8/14 = 8/14 - 8/14 = 0 \tag{2.21c}$$

$$u_2 \cdot \tilde{u}_2 = -1/2 \cdot 8/28 + 1 \cdot 8/7 = -4/28 + 32/28 = 1 \tag{2.21d}$$

and it is easily verified that any vector $v \in \mathbb{R}^2$ can indeed be written as

$$v = \langle v, \tilde{u}_1 \rangle \, u_1 + \langle v, \tilde{u}_2 \rangle \, u_2. \tag{2.22}$$

An estimate of the norm of the basis function coefficients for a biorthogonal basis is possible using the Bessel bound.

Proposition 2.17. *Let $\mathcal{H}$ be a Hilbert space, and let $(\varphi_i, \tilde{\varphi}_i)_{i=1}^k$ be a biorthogonal basis for $\mathcal{H}$. If $\{\varphi_i\}_{i=1}^k$ is a Bessel sequence with Bessel bound B, then for all $f \in \mathcal{H}$ it holds that*

$$\frac{1}{B}\|f\|^2 \leq \sum_{i=1}^k |\langle f, \tilde{\varphi}_i \rangle|^2.$$

Note that when $B = 0$ then the basis function coefficients are unbounded and reconstruction is not possible. Next, we will introduce some short-hand notation for the action the primal and dual basis functions by characterizing it using operators, and since these operators are well behaved, we will employ a proposition. In this context the discrete Lebesgue spaces ℓ_2^k, that were introduced in Example 2.19, naturally arise.

Proposition 2.18. *Let $\mathcal{H}$ be a Hilbert space, and let $(\varphi_i, \tilde{\varphi}_i)_{i=1}^k$ be a biorthogonal basis for $\mathcal{H}$. The **analysis operator** $\mathcal{A}$ is the bounded, linear operator*

$$\mathcal{A} = \mathcal{A}(\tilde{\varphi}) : \mathcal{H} \to \ell_2^k : f \mapsto \{c_i\}_{i=1}^k = \{\langle f, \tilde{\varphi}_i \rangle\}_{i=1}^k.$$

and its adjoint for $c = \{c_i\}_{i=1}^k \in \ell_2^k$ defined by

$$\langle \mathcal{A}f, c \rangle = \langle f, \mathcal{A}^*c \rangle$$

*is the bounded, linear **synthesis operator** $\mathcal{A}^*$ given by*

$$\mathcal{A}^* = \mathcal{A}^*(\tilde{\varphi}) : \ell_2^k \to \mathcal{H} : \{c_i\}_{i=0}^k \mapsto \sum_{i=0}^k c_i \tilde{\varphi}_i = g.$$

*The **reconstruction operator** $\mathcal{R}$ is the bounded, linear operator*

$$\mathcal{R} = \mathcal{R}(\varphi) : \ell_2^k \to \mathcal{H} : \{c_i\}_{i=0}^k \mapsto \sum_{i=0}^k c_i \varphi_i = f.$$

and $\mathcal{R}\mathcal{A} = I$, where I is the identity operator on $\mathcal{H}$.

In the definition of the synthesis operator by the adjoint of the analysis operator $\mathcal{A}$, the inner product on the left hand side is those in $\mathcal{H}$, and those on the right hand side is the inner product in ℓ_2^k. It is important to distinguish the synthesis operator $\mathcal{A}^*$ and the reconstruction operator $\mathcal{R}$, which employ the primary basis $\{\varphi_i\}_{i=1}^k$ and the dual basis $\{\tilde{\varphi}_i\}_{i=1}^k$, respectively, and the two only coincide when the basis has additional properties, as we will see in the following.

A corollary of the above proposition, which will be of relevance for computations in Hilbert spaces, is the following result.

Corollary 2.4. *Let $\mathcal{H}$ be a Hilbert space, and let $\{\varphi_i\}_{i=1}^k$ be a basis for $\mathcal{H}$. Then $\mathcal{H}$ is isomorphic to ℓ_2^k, with the isomorphism given by the analysis and reconstruction operators $\mathcal{A}(\tilde{\varphi})$ and $\mathcal{R}(\varphi)$, respectively.*

According to the above proposition, every biorthogonal basis establishes an isomorphism between $\mathcal{H}$ and ℓ_2^k. However, it is important to note that since the isomorphism depends on the basis $(\varphi_i, \tilde{\varphi}_i)_{i=1}^k$, some choices might be more suitable than others in specific contexts. With the above operators in hand, we can introduce another very useful tool to analyse and work with bases.

Proposition 2.19. *Let $\mathcal{H}$ be a Hilbert space with basis $\{\varphi_i\}_{i=1}^k$, and let $\mathcal{A} = \mathcal{A}(\varphi)$ and $\mathcal{A}^* = \mathcal{A}^*(\varphi)$ be the analysis and synthesis operators for the basis. Then the $\textbf{Gram matrix}$ or $\textbf{Gramian}$ $\mathcal{G} : \ell_2^k \to \ell_2^k$ is the bounded linear operator*

$$\mathcal{G} = \mathcal{G}(\varphi_i) = \mathcal{A}^* \mathcal{A} = \{\langle \varphi_i, \varphi_j \rangle\}_{i,j=1}^k .$$

Hence, the Gramian is a possibly infinite dimensional matrix whose $(i, j)^{\text{th}}$ entry is given by $\langle \varphi_i, \varphi_j \rangle$. Note that the Gramian is defined for a basis, without reference to its dual, although the properties of the Gramian for a primal and dual basis will certainly reflect this duality.

So far, we tacitly ignored a problem with biorthogonal Schauder bases: their convergence is not unconditional. We will now refine the concept in two ways to obtain bases for Hilbert spaces that overcome this limitation. The first refinement is an orthonormal basis, which generalizes the canonical basis $(e_1, \ldots, e_n)$ of Euclidean space $\mathbb{R}^n$ to arbitrary Hilbert spaces.

Definition 2.43. *Let $\mathcal{H}$ be a Hilbert space. An $\textbf{orthonormal basis}$ for $\mathcal{H}$ is a basis $\{\phi_i\}_{i=1}^k$ satisfying*

$$\langle \phi_i, \phi_j \rangle = \delta_{ij}.$$

A characterization of *all* orthonormal bases for a Hilbert space is provided by the following proposition, which employs the notion of a real unitary operator from Def. 2.38.

Proposition 2.20. *Let $\mathcal{H}$ be a Hilbert space and $\{\phi_i\}_{i=1}^k$ be an orthonormal basis for $\mathcal{H}$. If $U : \mathcal{H} \to \mathcal{H}$ is a real unitary operator, then the family of functions $\{\psi_i\}_{i=1}^k$ defined by*

$$\psi_i = U \phi_i$$

is again an orthonormal basis for $\mathcal{H}$, and all orthonormal basis can be generated from $\{\phi_i\}_{i=1}^k$ by a real unitary operator.

The importance of orthonormal bases lies largely in the following proposition and its consequences.

Proposition 2.21. *Let $\mathcal{H}$ be a Hilbert space and $\{\phi_i\}_{i=1}^k$ be an orthonormal basis for $\mathcal{H}$. Then $\{\phi_i\}_{i=1}^k$ is an unconditional basis for $\mathcal{H}$ and for any $f \in \mathcal{H}$ one has*

$$f = \sum_{i=1}^k \langle f, \phi_i \rangle \phi_i = \sum_{i=1}^k f_i \phi_i$$

The proposition shows that for an orthonormal basis the primary and dual basis functions coincide, and hence the same functions are employed to obtain the basis function coefficients and to reconstruct a function.

Remark 2.18. Every orthogonal basis can easily be normalized. Hence, for us orthogonal and orthonormal are synonymous, although, strictly speaking, an orthogonal basis is biorthogonal since primal and dual basis differ by a constant and do not coincide.

The following proposition summarizes additional valuable properties of orthonormal bases.

Proposition 2.22. *Let $\mathcal{H}$ be a Hilbert space and $\{\phi_i\}_{i=1}^k$ be an orthonormal basis for $\mathcal{H}$. Then for $f, g \in \mathcal{H}$ it holds*

i) *Parseval's identity, $\sum_{i=1}^k f_i = \sum_{i=1}^k \langle f, \phi_i \rangle = \|f\|^2$, and, conversely, Parseval's identity implies that $\{\phi_i\}$ is an orthonormal basis;*

ii) *$\langle f, g \rangle = \sum_{i=1}^k f_i \, g_i = \sum_{i=1}^k \langle f, \phi_i \rangle \langle g, \phi_i \rangle$.*

Remark 2.19. Many properties of orthonormal bases, such as Parseval's identity, are only well known for Fourier basis, and only believed to hold for such systems.[15] However, it is useful to keep in mind that these are satisfied for different bases whenever orthonormality of the basis functions holds.

A corollary of Proposition 2.21 is the following.

Corollary 2.5. *Let $\mathcal{H}$ be a Hilbert space and $\{\phi_i\}_{i=1}^\infty$ an orthonormal basis for the space. Then the synthesis operator $\mathcal{A}^*(\phi)$ and the reconstruction operator $\mathcal{R}(\phi)$ coincide.*

The above result together with Proposition 2.22 yields the following refinement of Corollary 2.4.

Corollary 2.6. . *Let $\mathcal{H}$ be a separable Hilbert space. Then $\mathcal{H}$ has an orthonormal basis $\{\phi_i\}_{i=1}^k$ and the space is isometrically isomorphic to ℓ_2^k with the isomorphism given by the analysis and reconstruction operators $\mathcal{A}(\phi)$ and $\mathcal{R}(\phi)$ with respect to $\{\phi_i\}_{i=1}^k$.*

For the next corollary of Proposition 2.22, which is of particular importance for applications, it is useful to keep the definition of convergence of a series representation in Def. 2.31 in mind.

Corollary 2.7. *Let $\mathcal{H}$ be a Hilbert space and $\{\phi_i\}_{i=1}^k$ be an orthonormal basis for $\mathcal{H}$. Then the **optimal m-term approximation***

$$\hat{f}_m = \min \left(\left\| f - \sum_{i \in I_m} \langle f, \phi_i \rangle \right\| \right) \in \mathcal{H}$$

---

[15]A point in case is (Ramamoorthi, Mahajan, and Belhumeur, "A First-Order Analysis of Lighting, Shading, and Shadows", p. 3).

is attained when the index set $I_m \subset \{1, \ldots, k\}$ with $\operatorname{card}(I_m) = m$ is chosen such that for any $j \notin I_m$ it holds that $f_i = \langle f, \phi_i \rangle > \langle f, \phi_j \rangle = f_j$ for all $i \in I_m$, that is

$$I_m = \{i \in \{1, \ldots, k\} \mid \langle f, \phi_i \rangle > \langle f, \phi_j \rangle, j > m\}.$$

The corollary states that the m largest basis function coefficients f_i of an element $f \in \mathcal{H}$ provides the optimal approximation with respect to the norm induced by the inner product. For biorthogonal bases, analogous result do not exists, and it is usually very difficult to find optimal approximations.

As Proposition 2.21 and Proposition 2.22 show, orthonormal bases have many very convenient properties. However, they can also be highly restrictive since orthonormality is a stringent requirement that makes it often difficult to enforce other useful properties, such as local support of the basis functions; in fact, at times it is difficult to obtain an orthonormal basis at all, despite the existence result in Theorem 2.6.[16] A requirement for a basis that is less stringent than orthonormality, but Which still guarantees unconditional convergence, is the following.

Definition 2.44. *Let $\mathcal{H}$ be a Hilbert space and let $\{\varphi_i\}_{i=0}^{k}$ be a basis for $\mathcal{H}$. Then the sequence $\{\varphi_i\}_{i=0}^{k}$ is a **Riesz basis** for $\mathcal{H}$ when there exists positive, real constant $A, B > 0$ such that for ever $f \in \mathcal{H}$ it holds that*

$$A\|f\|^2 \leq \sum_{i=1}^{k} \langle f, \varphi_i \rangle^2 \leq B\|f\|^2.$$

Intuitively, the constant A and B ensure that the analysis operator is invertible and that a well defined, that is bounded, reconstruction operator exists. This is made precise in the following theorem.

Theorem 2.8. *Let $\mathcal{H}$ be a Hilbert space and let $\{\varphi_i\}_{i=0}^{k}$ be a Riesz basis for $\mathcal{H}$. Then $\{\varphi_i\}_{i=0}^{k}$ is an unconditional basis for $\mathcal{H}$, and there exists a unique dual basis $\{\tilde{\varphi}_i\}_{i=0}^{k}$ satisfying the biorthogonality condition*

$$\langle \varphi_i, \tilde{\varphi}_j \rangle = \delta_{ij}$$

which is also an unconditional basis for $\mathcal{H}$. Every function $f \in \mathcal{H}$ admits the basis expansion

$$f = \sum_{i=1}^{k} \langle f, \tilde{\varphi}_i \rangle \, \varphi_i = \sum_{i=1}^{k} \langle f, \varphi_i \rangle \, \tilde{\varphi}_i.$$

A Riesz basis is hence an conditional basis for $\mathcal{H}$ where the primary and dual basis functions are fully equivalent. Unless mentioned otherwise, in the following 'basis' will always refer to a Riesz basis.

An alternative characterization of a Riesz basis, which is an extension and generalization of Proposition 2.20, is the following result.

---

[16]See for example the discussion in (Lessig and Fiume, "SOHO: Orthogonal and Symmetric Haar Wavelets on the Sphere").

Proposition 2.23. *Let $\mathcal{H}$ be a Hilbert space, $\{\phi_i\}_{i=0}^k$ an orthonormal basis for $\mathcal{H}$, and $\{\varphi_i\}_{i=1}^k$ a Riesz basis pair for $\mathcal{H}$. Then there exists a bounded, bijective operator $U : \mathcal{H} \to \mathcal{H}$ such that*

$$\varphi_i = U\,\phi_i,$$

and every such operator defines a Riesz basis.

With a Riesz basis, one loses the ability to easily find the optimal m-term approximation of a function, which for an orthonormal basis was possible with Proposition 2.7. Nonetheless, the unconditional convergence and the flexibility which they afford make them highly useful in practice. In fact, most "modern bases", such as second and third generation wavelets, are Riesz bases. From Parseval's identity in Proposition 2.22 it is also apparent that an orthonormal basis is also Riesz basis with constants $A = B = 1$.

So far we considered the existence of bases and the different properties they have, but we did not discuss what bases are suitable in applications and how these can be obtained. In general, this is a daunting question. However, when a problem involves a self-adjoint, compact operator, then an answer is provided by the following result, which concludes our discussions of such operators from Theorem 2.6.

Proposition 2.24. *Let $\mathcal{H}$ be a Hilbert space, and let $A : \mathcal{H} \to \mathcal{H}$ be a compact, self-adjoint operator on $\mathcal{H}$. Then any function in the range $\mathrm{ran}\,(A)$ of A can be represented using the eigenfunctions u_i corresponding to the nonzero eigenvalues λ_i of A, and the u_i can be chosen to form an orthonormal basis for the space. Moreover, the image of a function $f \in \mathcal{H}$ under A can be represented as*

$$Af = \sum_{i=1}^k \lambda_i\, f_i\, u_i$$

where the $f_i = \langle f, u_i \rangle$ are the basis function coefficients of f with respect to the basis $\{u_i\}_{i=1}^k$ for $\mathrm{ran}\,(A)$, and the index set runs over all nonzero eigenvalues of A.

The last assertion follows from a simple but instructive calculation. By definition we have

$$Af = \sum_{i=1}^k \langle Af, u_i \rangle\, u_i \tag{2.23a}$$

and since A is self-adjoint we obtain

$$Af = \sum_{i=1}^k \langle f, A\,u_i \rangle\, u_i. \tag{2.23b}$$

With the definition of the eigenvalues and eigenvectors from Def. 2.41 we have

$$Af = \sum_{i=1}^k \langle f, \lambda_i\,u_i \rangle\, u_i \tag{2.23c}$$

and exploiting linearity yields

$$Af = \sum_{i=1}^{k} \lambda_i \langle f, u_i \rangle u_i \qquad (2.23d)$$

which immediately implies the result by the defintion of the basis function coefficients f_i. The above proposition shows that using the eigenfunctions of A as basis "diagonalizes" the operator, and its application to $f \in \mathcal{H}$ is trivial in that it only requires a multiplication of the basis function coefficients with the scalar eigenvalues λ_i.

Remark 2.20. When A is an integral operator with kernel $k(x, y)$, and $\mathcal{H}$ is a closed subspace of $L_2(X)$ such that the inner product is given by integration, then it follows from Eq. 2.23c that the kernel $k(x, y)$ has the representation

$$k(x, y) = \sum_{i=1}^{k} \lambda_i \, u_i(x) \, u_i(y).$$

Proposition 2.24 can be extremely useful when one encounters a compact and self-adjoint operator. When self-adjointness is not satisfied then the singular value decomposition can provide a remedy.[17]

Proposition 2.25. *Let $\mathcal{H}$ be a Hilbert space and $A : \mathcal{H} \to \mathcal{H}$ be a compact intergal operator, not necessarily self-adjoint, with adjoint operator A^*. Then the left $L = A^*A$ and right $R = AA^*$ iterates are integral operators and their spectrum consists only of eigenvalues and possibly zero, which is also the only possible accumulation point. Moreover, the eigenvalues of L and R coincide and are the **singular values** σ_i **of** A, and one has*

$$Lu_i = \sigma_i u_i \qquad Rv_i = \sigma_i v_i$$

*where $\{u_i\}_{i=1}^{k}$ and $\{v_i\}_{i=1}^{k}$ are the eigenvectors of L and R, respectively. The operator A admits the **singular value decomposition***

$$Au_i = \lambda_i v_i$$

*and $\{u_i\}_{i=1}^{k}$ and $\{v_i\}_{i=1}^{k}$ are the **left and right singular vectors of** A. A function $h \in \mathrm{ran}(A)$ in the range of A can be represented as*

$$h = \sum_{i=1}^{k} \langle h, v_i \rangle v_i$$

while a function $g \in \mathrm{ran}(A^*)$ in the range of the adjoint A^* can be written as

$$g = \sum_{i=1}^{k} \langle g, u_i \rangle u_i.$$

---

[17]The following material is from (Stakgold and Holst, *Green's Functions and Boundary Value Problems*, Exercise 6.3.2). For a more rigorous treatment see also (Lax, *Functional Analysis*, Chapter 30).

Moreover, for $f \in \mathcal{H}$ the action of A is given by

$$Af = \sum_{i=1}^{k} \sigma_i \, f_i \, v_i$$

where the $f_i = \langle f, u_i \rangle$ are the basis function coefficients of f with respect to $\{u_i\}_{i=1}^{k}$.

The last statement follows from a calculation analogous to those in Eq. 2.23. Note that the range of A^* is a subset of the domain of A, so that the $\{u_i\}_{i=1}^{k}$ can also be interpreted as providing a representation for the pre-image of A. Proposition 2.24 for a self-adjoint operator can be considered as a special case of the singular value decomposition where the left and right singular functions u_i and v_i coincide.

2.2.2.4 Hilbert Space Frames[18]

At the beginning of our discussion of linear spaces, we introduced in Proposition 2.3 an arbitrary collection $\{x_i\}_{i=1}^{k}$ of elements $x_i \in \mathrm{U}$ to represent all $y \in \mathrm{U}$ as linear combinations of the form

$$y = \sum_{i=1}^{m} a_i \, x_i. \tag{2.24}$$

Only later on, when we introduced the concept of a Schauder basis, we required our spanning set to be linearly independent. One can hence ask the question, which price we paid by restricting ourselves to linearly independent sets, and if it can be useful to loosen the restrictions on the set of elements we employ to represent arbitrary vectors. Let us begin by considering an example.

Example 2.22. Any vector $v \in \mathbb{R}^2$ can be written as a linear combination of the canonical basis vectors e_1 and e_2 for $\mathbb{R}^2$, that is

$$v = \langle v, e_1 \rangle \, e_1 + \langle v, e_2 \rangle \, e_2 = v_1 \, e_1 + v_2 \, e_2 = \sum_{i=1}^{2} v_i \, e_i. \tag{2.25}$$

However, can v also be written as a linear combination when additionally a third vector such that

$$e_3 = e_1 - e_2 \tag{2.26}$$

is employed, cf. Fig. 2.2? Introducing the vector space identity

$$0 = (v_1 - v_1) \, e_3 = (v_1 - v_1)(e_1 - e_2) \tag{2.27}$$

---

[18]The material in this section is largely drawn from the excellent introduction to the subject in (Kovacevic and Chebira, "Life Beyond Bases: The Advent of Frames (Part I)"), with some technical results from (Christensen, *An Introduction to Frames and Riesz Bases*, Chapter 5). Frames were first introduced into the literature in a landmark paper by Duffin and Schaeffer ("A Class of Nonharmonic Fourier Series"), although the idea remained obscure until the rise of wavelets in the 1980s. Interestingly, Duffin and Schaeffer employed the concept for a problem very similar to the ones we have in mind and which will be discussed in Chapter 4.
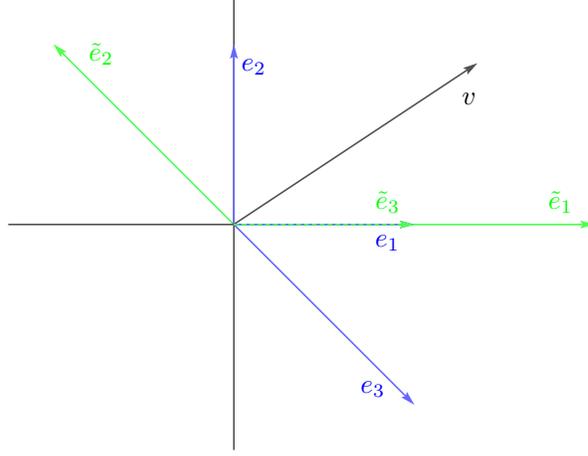
Figure 2.2: Hilbert space frame for $\mathbb{R}^2$ defined by three vectors (e_1, e_2, e_3). Any vector v in $\mathbb{R}^2$ can be represented using the three "basis vectors" (e_1, e_2, e_3).

and adding it to Eq. 2.25 we obtain

$$v = v_1\, e_1 + v_2\, e_2 + 0 \tag{2.28a}$$

$$= \langle v, e_1 \rangle\, e_1 + \langle v, e_2 \rangle\, e_2 + (\langle v, e_1 \rangle - \langle v, e_1 \rangle)(e_1 - e_2). \tag{2.28b}$$

Exploiting linearity and re-arranging terms, in a manner that for the moment probably appears slightly artifical, we obtain

$$v = \langle v, e_1 \rangle\, e_1 + \langle v, e_1 \rangle\, e_1 + \langle v, e_2 \rangle\, e_2 - \langle v, e_1 \rangle\, e_2 - \langle v, e_1 \rangle\,(e_1 - e_2) \tag{2.28c}$$

$$= 2\langle v, e_1 \rangle\, e_1 + (\langle v, e_2 \rangle - \langle v, e_1 \rangle)\, e_2 - \langle v, e_1 \rangle(e_1 - e_2) \tag{2.28d}$$

$$= \langle v, 2e_1 \rangle\, e_1 + \langle v, e_2 - e_1 \rangle\, e_2 - \langle v, e_1 \rangle\, e_3. \tag{2.28e}$$

The last equation shows that an arbitrary vector $v \in \mathbb{R}^2$ can indeed be expanded using the three "basis vectors" $\{e_1, e_2, e_3\}$. Moreover, with our foregoing discussion on biorthogonal bases in mind, it is natural to identify the "dual basis vectors"

$$\tilde{e}_1 = 2e_1 \qquad \tilde{e}_2 = e_2 - e_1 \qquad \tilde{e}_3 = -e_1. \tag{2.29}$$

We hence obtain the representation

$$v = \langle v, \tilde{e}_1 \rangle\, e_1 + \langle v, \tilde{e}_2 \rangle\, e_2 + \langle v, \tilde{e}_3 \rangle\, e_1 \tag{2.30a}$$

$$= \tilde{v}_1\, e_1 + \tilde{v}_2\, e_2 + \tilde{v}_3\, e_3 \tag{2.30b}$$

$$= \sum_{i=1}^{3} v_i\, e_i \tag{2.30c}$$

which looks formally like a basis expansion. Interestingly, when, instead of Eq. 2.27, we start with the vector space identity

$$0 = (v_2 - v_2)\, e_3 = (v_2 - v_2)(e_1 - e_2), \tag{2.31}$$

we obtain, using a derivation analogous to Eq. 2.28, that

$$v = \langle v, e_1 + e_2 \rangle\, e_1 + \langle v, 2e_2 \rangle\, e_2 - \langle v, e_2 \rangle\, e_2. \tag{2.32}$$

Comparing to Eq. 2.28e, we see that the "dual basis vectors" $\tilde{e}_i$, and hence also the expansion coefficients obtained by the inner product $\langle v, \tilde{e}_i \rangle$, are no longer unique when the overcomplete "basis" $\{e_1, e_2, e_3\}$ is employed.

Given the non-uniqueness of "dual basis functions" and expansion coefficients, one might ask why consider overcomplete representations with "basis vectors" than necessary should be considered. A simple answer is the greater flexibility afforded by such representations where one foregoes with linear independence. For example, one can design overcomplete expansions that can adapt to all possible features a signal might have, and then employ for reconstruction only those "basis functions" that correspond to feature in a particular signal, very much in the spirit of the optimal m-term approximation using orthonormal bases that was discussed in Corollary 2.7. Additionally, overcomplete representations are redundant, which is useful in "noisy" applications where information might be lost, for example when it is transported over a real-world channel, or when only partial information is available. In fact, our main motivation for the use of frames in Chapter 4 is the resilience of frames to partial information about the nature of our signals.[19]

Remark 2.21. In infinite dimensions, care is required when overcompleteness is to be defined, since there linear independence loses its classical meaning, and for example overcomplete representations with linearly independent spanning elements exist. This leads to the notion of ω-independence, which is however beyond the scope of our treatment.[20]

With these words of caution, and the above motivation in mind, we will now formalize the notion of an overcomplete representation.

Definition 2.45. *Let $\mathcal{H}$ be a Hilbert space and $\{\psi_i\}_{i=1}^m$ be a sequence in $\mathcal{H}$, with n being a natural number including infinity. Then $\{\psi_i\}_{i=1}^m$ is a Hilbert space **frame** if there are real, positive constants $A, B > 0$ such that for every $f \in \mathcal{H}$ it holds*

$$A\|f\|^2 \leq \sum_{i=1}^m |\langle f, \psi_i \rangle|^2 \leq B\|f\|^2.$$

---

[19]See for example (Kovacevic and Chebira, "Life Beyond Bases: The Advent of Frames (Part I)"; Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*) for a more detailed discussion of the practical advantages of frames.

[20]Christensen, *An Introduction to Frames and Riesz Bases*, Chapter 3.

*The constants A and B are then known as **lower frame bound** and **upper frame bound**, respectively, and the supremum and infinum are the optimal frame bounds.*

Note that there is no constraint on the cardinality of $\{\psi_i\}_{i=1}^m$. However, when $\mathrm{card}\,(\{\psi_i\}_{i=1}^m) < \dim(\mathcal{H})$ then A is necessarily zero, since a k-dimensional space cannot be spanned with less than k vectors, and hence the sequence $\{\psi_i\}_{i=1}^m$ is no longer a frame. The upper frame bound B ensures that the analysis operator for a frame is bounded, and hence a reconstruction from the frame coefficients $\langle f, \psi_i \rangle$ is possible, cf. Proposition 2.17. It is also important to note that the frame bounds are not unique, although we will usually employ the optimal frame bounds. Def. 2.45 should be compared to the definition of a Riesz basis in Def. 2.44, see also Example 2.24 below.

Remark 2.22. In the mathematics literature, 'frame' can refer to a variety of concepts, such as Hilbert space frame, orthonormal frame, moving frame, frame bundle, projective frame, and hence one has to be careful which notion is referred to.

Before we continue, let us given some more examples of frames.

Example 2.23. An orthonormal basis $\{\phi_i\}_{i=1}^k$ for a Hilbert space $\mathcal{H}$ is a frame with $k = n = \dim(\mathcal{H})$. The frame bounds are unity in this case, that is $A = B = 1$.

Example 2.24. A biorthogonal Riesz basis $\{\varphi_i\}_{i=1}^k$ for a Hilbert space $\mathcal{H}$ is a frame with $k = n = \dim(\mathcal{H})$.

Hence, we could have introduced Riesz bases as frames without redundancy.[21] The above examples also shows that the frame bounds can be employed to characterize different kinds of frames. This is formalized in the following definition.

Definition 2.46. *Let $\mathcal{H}$ be a Hilbert space and $\{\psi_i\}_{i=1}^m$ be a frame for $\mathcal{H}$. Then $\{\psi_i\}_{i=1}^m$ is*

i) tight, when the frame bounds can be chosen such that $A = B$;

ii) Parseval tight, when the frame bounds can be chosen such that $A = B = 1$;

iii) uniform, when $\|\psi_i\| = 1$.

It has to be noted that the nomenclature in the literature on Hilbert space frames is far from uniform, which can lead to much confusion. Obviously, the above characteristics are not exclusive, and an orthonormal basis is for example a uniform, Parsevel tight frame. As a generalization of Proposition 2.20 for orthonormal bases, we have the following result which shows that the above frame properties are stable under unitary mappings.

---

[21] In fact, this approach is sometimes taken in the literature. See for example (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Chapter 5).

Proposition 2.26. *Let $\mathcal{H}$ be a Hilbert space and $\{\psi_i\}_{i=1}^m$ be a frame for $\mathcal{H}$ with frame bounds A, B, and let $U : \mathcal{H} \to \mathcal{H}$ be a unitary operator acting on $\mathcal{H}$. Then $\{U\psi_i\}_{i=1}^m$ is again a frame for $\mathcal{H}$ with frame bounds A, B.*

A characteristic of finite dimensional frames is the following.

Definition 2.47. *Let $\mathcal{H}$ be finite Hilbert space of dimension k, and let $\{\psi_i\}_{i=1}^m$ be a frame for $\mathcal{H}$. Then the **redundancy** of $\{\psi_i\}_{i=1}^m$ is*

$$R(\psi) = \frac{m}{k}.$$

In analogy to Theorem 2.8 for Riesz bases, we have the following central result which justifies the relevancy of frames for applications.

Theorem 2.9. *Let $\mathcal{H}$ be a Hilbert space and $\{\psi_i\}_{i=1}^k$ be a frame for $\mathcal{H}$. Then there exists a **dual frame** $\{\tilde{\psi}_i\}_{i=1}^k$ formed by elements $\psi_i \in \mathcal{H}$, and every $f \in \mathcal{H}$ admits a frame expansion*

$$f = \sum_{i=1}^k \langle f, \tilde{\psi}_i \rangle \, \psi_i.$$

Moreover, the dual frame $\{\tilde{\psi}_i\}_{i=1}^k$ forms a frame for $\mathcal{H}$ with lower frame bound $\tilde{A} = B^{-1}$ and upper frame bound $\tilde{B} = A^{-1}$.

We emphasize again that the dual frame is not unique. However, we will introduce a canonical dual shortly, and we will consider the trade-off between different duals also in the next section when we restrict us to the finite dimensional case. For a tight frame, we have the following result for the expansion of a function.

Proposition 2.27. *Let $\mathcal{H}$ be a Hilbert space and $\{\psi_i\}_{i=1}^k$ be a tight frame for $\mathcal{H}$ with frame constant A. Then every $f \in \mathcal{H}$ admits the frame expansion*

$$f = \frac{1}{A} \sum_{i=1}^k \langle f, \psi_i \rangle \, \psi_i.$$

With the above proposition, we can think of a tight frame as the overcomplete or redundant analogue of an orthonormal basis. However, the Parseval identity in Proposition 2.22 is in general lost, although the scaling of the norm is easily characterized, as the following proposition shows.

Proposition 2.28. *Let $\mathcal{H}$ be a Hilbert space and $\{\psi_i\}_{i=1}^k$ be a tight frame with frame constant A, and let $f \in \mathcal{H}$. Then*

$$\sum_{i=1}^k |\langle f, \psi_i \rangle|^2 = A\|f\|^2.$$

If $\{\psi_i\}_{i=1}^k$ is a uniform tight frame with $\|\psi_i\| = 1$, then $A = R(\psi)$ and the frame bound A coincides with the redundancy $R(\psi)$ of the frame.
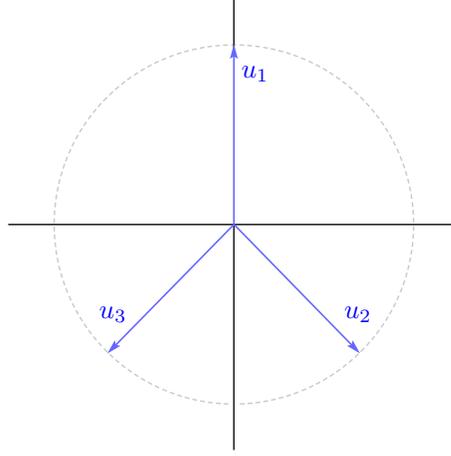
Figure 2.3: Mercedes-Benz frame.

The above proposition shows that when $A = B = 1$, and $\{\psi_i\}_{i=1}^m$ is a Parseval tight frame, then the expansion is in fact norm preserving. This is consistent with our previous considerations since when $A = B = 1$ and the frame functions are normalized, then $\{\psi_i\}_{i=1}^m$ is necessarily an orthonormal basis.

Example 2.25. The classical example of a frame for $\mathbb{R}^2$ is the Mercedes-Benz frame shown in Fig. 2.3. The frame vectors are given by

$$u_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad u_2 = \begin{pmatrix} \sqrt{3}/2 \\ -1/2 \end{pmatrix} \qquad u_3 = \begin{pmatrix} -\sqrt{3}/2 \\ -1/2 \end{pmatrix} \tag{2.33}$$

and these form a uniform, tight frame with frame constant $A = 2/3$. Hence, the frame inequality for the Mercedes-Benz frame is

$$\frac{2}{3}\|v\|^2 \le \sum_{i=1}^{3} |\langle v, u_i \rangle|^2 \le \frac{2}{3}\|v\|^2 \tag{2.34}$$

and we thus have

$$\sum_{i=1}^{3} |\langle v, u_i \rangle|^2 = \frac{2}{3}\|v\|^2 \tag{2.35}$$

which is just Proposition 2.28. The expansion of an arbitary vector $v \in \mathbb{R}^2$ in the Mercedes-Benz frame is given by

$$v = \frac{3}{2} \sum_{i=1}^{3} \langle v, u_i \rangle u_i. \tag{2.36}$$

Remark 2.23. Similar to the situation for orthogonal bases, the construction of tight frames is by no means simple, and even in the finite dimensional case challenging. An interesting characterization of tight frames employs potential theory and a frame force F_F given by

$$F_F(\psi) = 2\langle\psi_i, \psi_j\rangle(\psi_i - \psi_j)$$

which "repulses" frame vectors ψ_i and ψ_j that are "too close" and "attracts" vectors that are "too far" apart. Uniform tight frames are then m-point equilibrium states of points on the k-sphere under the potential generated by the frame force F_F, where, as usual, k is the dimension of the space and m the number of frame vectors.[22] For example, the Mercedes-Benz frame, which was introduced in the foregoing example, is the minimizer of the frame force for three points on the unit circle S^1 in the plane $\mathbb{R}^2$. Unfortunately, finding equilibrium configurations under a potential even on the 2-sphere is very hard, and it is is an even harder problem in higher dimensions.[23] This makes the concept of a frame force non-constructive in a classical sense, although numerical techniques can typically be employed to minimize the frame potential.

The frame operator, which is introduced next, is an important tool to characterize frames. We refer to Def. 2.18 for the definition of the analysis operator $\mathcal{A}$, which, with obvious modifications, carries over to the case of a frame.

Definition 2.48. *Let $\mathcal{H}$ be a Hilbert space, and let $\{\psi_i\}_{i=1}^k$ be a frame for $\mathcal{H}$ with analysis operator $\mathcal{A} = \mathcal{A}(\psi)$. Then the **frame operator** $\mathcal{S}$ is*

$$\mathcal{S} = \mathcal{S}(\psi_i) = \mathcal{A}\mathcal{A}^*.$$

The frame operator $\mathcal{S}$ is the "dual" to the Gramian $\mathcal{G}$ that was defined as $\mathcal{G} = \mathcal{A}^*\mathcal{A}$, cf. Def. 2.19. Important properties of the frame operator $\mathcal{S}$ and a first application is provided in the following proposition.

Proposition 2.29. *Let $\mathcal{H}$ be a Hilbert space, and let $\{\psi_i\}_{i=1}^k$ be a frame for $\mathcal{H}$ with frame operator $\mathcal{S} = \mathcal{S}(\psi)$. Then $\mathcal{S}$ is bounded, invertible, self-adjoint, and positive, and it defines **canonical dual frame functions** $\tilde{\psi}_i \in \mathcal{H}$ by*

$$\tilde{\psi}_i = \mathcal{S}^{-1}\psi_i.$$

Remark 2.24. For a tight frame, the action of the frame operator can be understood as the projection onto the space spanned by the frame vectors. More precisely let $\bar{\mathcal{H}}$ be a closed subspace of $\mathcal{H}$ so that the inner product on the spaces is identical. Then for $f \in \mathcal{H}$ we have $\mathcal{S}f = \bar{f}$, where $\bar{f}$ is the orthogonal projection from $\mathcal{H}$ onto $\bar{\mathcal{H}}$, cf. Remark 2.10.

---

[22]Benedetto and Fickus, "Finite Normalized Tight Frames"; Fickus et al., "Convolutional frames and the frame potential".

[23]See for example (Saff and Kuijlaars, "Distributing many points on a sphere"; Armentano, Beltrán, and Shub, "Minimizing the discrete logarithmic energy on the sphere: The role of random polynomials") and references therein.

With the frame operator $\mathcal{S}$, we can characterize the frame bounds also using the eigenvalues of $\mathcal{S}$, which are for our purposes in the following result well defined by Proposition 2.6 and the fact that every finite operator is compact.

Proposition 2.30. *Let $\mathcal{H}$ be a finite dimensional Hilbert space with $\dim(\mathcal{H}) = n$, and $\{\psi_i\}_{i=1}^{m}$ be a frame for $\mathcal{H}$ with frame operator $\mathcal{S}$. Then the smallest eigenvalue λ_1 of $\mathcal{S}$ is the optimal lower frame bounds, while the largest eigenvalue λ_n of $\mathcal{S}$ is the optimal upper frame bound.*

This is consistent with Proposition 2.6 since the kernel is empty if and only if the smallest eigenvalue is nonzero, and only in this case is $\mathcal{S}$ invertible, which by Proposition 2.29 is required to obtain the dual frame functions for reconstruction. By Proposition 2.25, the result could have been stated equivalently for the Gramian.

We only presented the basic notions of frames that are relevant for our purposes, and we have to refer to the literature for a more detailed discussion of this very active area of research, where many open questions still exist even in the finite dimensional case.[24]

Remark 2.25. The foregoing discussion shows that Hilbert space frames are generalization of Riesz bases, or more precisely that they provide a more general theory for such expansions. Unless mentioned otherwise, 'basis' will in the following refer to a possibly overcomplete basis, and hence include frames, and sometimes we will also just speak of a representation. Obviously, we will be careful when the difference is important.

2.2.3 Reproducing Kernel Hilbert Spaces[25]

We began this section with "plain" vector spaces, later added a norm to obtain normed spaces, completing such spaces provided us with Banach spaces, and then adding an inner product gave us Hilbert spaces. In the following, we will furnish our spaces with even more structure, which, as we will see in Chapter 4, makes them particularly well suited for numerical computations. In this section, we will always assume that our Hilbert space is a function space defined over a set X, such as $X = [a, b] \subset \mathbb{R}$.

---

[24]See for example (Kovacevic, Dragotti, and Goyal, "Filter bank frame expansions with erasures"; Benedetto and Fickus, "Finite Normalized Tight Frames"; Vale and Waldron, "Tight Frames and Their Symmetries"; Fickus et al., "Convolutional frames and the frame potential"; Gavruta, "On some identities and inequalities for frames in Hilbert spaces"; Bodmann and Casazza, "The road to equal-norm Parseval frames") and references therein.

[25]Reproducing kernel Hilbert spaces were considered sporadically in the first half of the 20th century, with important contributions for example by Mercer ("Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations"), Moore ("On Properly Positive Hermitian Matrices"), and Bergmann ("Über die Entwicklung der harmonischen Funktionen der Ebene und des Raumes nach Orthogonalfunktionen"), and they were first studied systematically in a landmark paper by Aronszajn ("Theory of Reproducing Kernels"). More recent treatments of the theory can for example be found in (Meschkowski, *Hilbertsche Räume mit Kernfunktion*; Saitoh, *Integral Transforms, Reproducing Kernels and their Applications*), although it is fair to say that currently no adequate modern monograph on the theory exists.

Definition 2.49. *Let $\mathcal{H}(X)$ be a Hilbert space over a set X. Then the **point evaluation functional** $\delta_y \in \mathcal{H}(X)$ **at** $y \in X$ is the functional $\delta_y : \mathcal{H}(X) \to \mathbb{R}$ that for every $f \in \mathcal{H}(X)$ satisfies*

$$\delta_y(f) = f(y).$$

What makes the point evaluation functional δ_y interesting is its reproducing or sifting property, which "reproduced" the value of the function f at the location $y \in X$. In general, the point evaluation functional is not continuous, and it is hence not an element in the continuous dual space $\mathcal{H}^*$. As is well known, it is then usually a distribution in the sense of Schwartz and known as Dirac delta or, with abuse of notation, as Dirac delta function. Of interest to us is the restricted class of Hilbert space where the point evaluation functional is bounded.

Definition 2.50. *Let $\mathcal{H}(X)$ be a Hilbert space over a set X. Then $\mathcal{H}(X)$ is a **reproducing kernel Hilbert space** $\mathcal{H}_k(X)$ when the point evaluation functional δ_y for $\mathcal{H}(X)$ is continuous for all $y \in X$.*

The definition poses the question for which Hilbert spaces the point evaluation functional is bounded. Unfortunately, there is no characterization that enables to easily answer the question, although the examples we will consider later will show that reproducing kernel Hilbert space are formed by "well behaved" functions. However, as a consequence of Proposition 2.8 we have the following result which characterizes the reproducing kernel Hilbert spaces of interest for numerical computations, and hence also of most interest for us.

Proposition 2.31. *Let $\mathcal{H}$ be a Hilbert space that is finite dimensional. Then $\mathcal{H}$ is a reproducing kernel Hilbert space.*

The following proposition, which is central for the theory of reproducing kernel Hilbert spaces, is an immediate consequence of Def. 2.50 and the Riesz representation theorem in Theorem 2.5.

Proposition 2.32. *Let $\mathcal{H}_k(X)$ be a reproducing kernel Hilbert space defined over a set X. Then the **reproducing kernel** k_y **for** $\mathcal{H}_k(X)$ is the unique element in $\mathcal{H}_k(X)$ such that for every $f \in \mathcal{H}_k(X)$ it holds*

$$\delta_y(f) = \langle k_y(x), f \rangle = f(y).$$

In the above proposition, the Riesz representation theorem enabled to identify the point evaluation functional δ_y with a function

$$k_y(x) = k(y, x) \in \mathcal{H}_k(X) \tag{2.37}$$

which, through the inner product of $\mathcal{H}_k(X)$, also has the reproducing property, see Fig. 2.4. It is also important to keep in mind that the reproducing kernel is
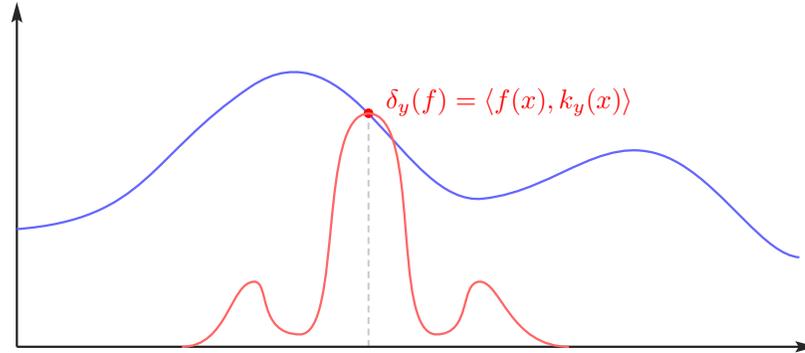
Figure 2.4: In a reproducing kernel Hilbert space $\mathcal{H}_k(X)$ the point evaluation functional δ_y can be identified with a function $k_y(x) \in \mathcal{H}_k(X)$ whose action is described by the inner product.

unique for a Hilbert space, and it can hence be used to characterize a space.[26] Important properties of the reproducing kernel are summarized in the following proposition.

Proposition 2.33. *Let $\mathcal{H}_k(X)$ be a reproducing kernel Hilbert space defined over a set X with reproducing kernel $k_y(x) = k(y,x)$, and let $x,y \in X$. Then $k_y(x)$ satisfies*

   i) positivity, $k(x,x) > 0$;

  ii) symmetry, $k(x,y) = k(y,x)$;

 iii) $|k(x,y)|^2 < k(x,x)\,k(y,y)$;

 iv) $|f(y)| \le K(y,y)^{1/2}\,\|f\|$;

  v) for all functions $f \in \mathcal{H}_k(X)$ satisfying $f(y) = 1$, $k(y,y)^{-1}k_y(x)$ is the function with this property with minimal norm over $\mathcal{H}_k(X)$;

 vi) for all functions $\|f\| \le 1$ in the unit ball in $\mathcal{H}_k(X)$, the absolute value $|f(y)|$ at any $y \in X$ satisfies $|f(y)| \le \|k(x,y)\|^{-1}k(x,y)$.

The last two properties are important in approximation theory in reproducing kernel Hilbert spaces and related areas,[27] since it often allows to establish a result for the reproducing kernel that then provides a bound for all other functions in

---

[26]The use of reproducing kernels to characterize function spaces is often encountered in representation theory, see for example (Helgason, *Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators and Spherical Functions*), and the kernels are then often known as zonal spherical functions, terminology which originates in the theory of spherical harmonics under the action of SO(3), cf. Chapter 2.2.5.2.

[27]See for example (Novak and Woźniakowski, *Tractability of Multivariate Problems: Standard Information for Functionals*) and references therein.

the space. That reproducing kernel Hilbert spaces are "well behaved" spaces is shown by the following result, see Remark 2.30 for the notions of convergence employed here.

Proposition 2.34. *Let $\mathcal{H}_k(X)$ be a reproducing kernel Hilbert space over a set X with reproducing kernel $k_y(x)$, which, as a function, is bounded on X. Then a sequence $\{f_i\}_{i=1}^{\infty} \to f$ which converges in the norm of $\mathcal{H}_k$ converges pointwise for every $x \in X$, and, moreover, the pointwise convergence is uniform.*

Intuitively, the existence of a reproducing kernel is a consequence of the uniform pointwise convergence in the above proposition that ensures that pointwise limits are well defined. An important property of reproducing kernel Hilbert spaces, in particular when one has a sequence of such spaces, is the following.

Proposition 2.35. *Let $\mathcal{H}_k(X)$ be a reproducing kernel Hilbert space over a set X with reproducing kernel $k_y(x)$, and let $\bar{\mathcal{H}}_k(X)$ be a closed linear subspace of $\mathcal{H}_k(X)$. Then $\bar{\mathcal{H}}_k(X)$ is a reproducing kernel Hilbert space with kernel $\bar{k}_y(x)$, and the projection $\bar{f} \in \bar{\mathcal{H}}_k(X)$ of any $f \in \mathcal{H}_k(X)$ onto the subspace $\bar{\mathcal{H}}_k(X)$ is given by*

$$\bar{f} = \langle f, \bar{k}(x, y) \rangle.$$

So far we merely posited the existence of a reproducing kernel. The next proposition provides a concrete recipe for the construction of a reproducing kernel in a separable Hilbert space.

Proposition 2.36. *Let $\mathcal{H}_k$ be a separable reproducing kernel Hilbert space of dimension n, with n possibly being infinity, and let $\{\phi_i\}_{i=1}^{n}$ be an orthonormal basis for $\mathcal{H}_k(X)$. Then the reproducing kernel $k_y(x)$ for $\mathcal{H}_k(X)$ is given by*

$$k_y(x) = k(y, x) = \sum_{i=1}^{n} \phi_i(y)\,\phi_i(x).$$

The above result, which is a form of Mercer's theorem, is of central importance because it enables us to construct a reproducing kernel for a Hilbert space.

Remark 2.26. When only a biorthogonal basis pair $(\psi_i, \tilde{\psi}_i)_{i=1}^{n}$ for $\mathcal{H}_k(X)$ is available, then it takes the form

$$k_y(x) = k(y, x) = \sum_{i=1}^{n} \psi_i(y)\,\tilde{\psi}_i(x). \tag{2.38}$$

With the definition of the reproducing kernel in Proposition 2.36, it is useful to consider the foregoing propositions again.

Remark 2.27. Much intuition for reproducing kernels in separable Hilbert spaces can be gained by expanding its definition in Proposition 2.36. We then have

$$f(y) = \langle f(x), k(x, y) \rangle \tag{2.39a}$$

$$= \left\langle f(x), \left(\sum_{i=1}^{n} \phi_i(x)\,\phi_i(y) \right) \right\rangle \tag{2.39b}$$

and using linearity yields

$$= \sum_{i=1}^{n} \langle f(x), \phi_i(x) \rangle \; \phi_i(y) \tag{2.39c}$$

$$= \sum_{i=1}^{n} f_i\,\phi_i(y), \tag{2.39d}$$

which is the usual formula for the reconstruction of $f \in H_k(X)$ from its basis function coefficients f_i. Reproducing kernels hence naturally arise as pointwise projection operators for Hilbert spaces, which was just Proposition 2.35.[28] The reproducing property for functions in $\mathcal{H}_k(X)$ in the definition of the reproducing kernel can thus be interpreted as the well known idempotence of projections operators.

Remark 2.28. An alternative interpretation of the formula in Proposition 2.36 that is often useful is to consider

$$k(y, x) = \sum_{i=1}^{n} \phi_i(y)\,\phi_i(x)$$

as the basis expansion of the kernel $k(y, x)$ with respect to $\{\phi_i\}_{i=1}^{k}$ with basis function coefficients $\{\phi_i(y)\}_{i=1}^{k}$.

In the following, we will consider some classical examples of reproducing kernel Hilbert spaces. However, let us begin with a counter-example.

Example 2.26. The space $L_2(X)$ is *not* a reproducing kernel Hilbert space. The point evaluation functional for $L_2(X)$ is the Dirac delta, and it is a distribution in the sense of Schwartz.

Example 2.27. By the Sobolev embedding theorem, for $s > d/2$ the Hilbert-Sobolev spaces $H^s(\mathbb{R}^d)$ are reproducing kernel Hilbert spaces, cf. Example 2.18.[29]

Example 2.28. The spaces spanned by classical spline functions are reproducing kernel Hilbert space.[30]

---

[28]Note that when a reproducing kernel acts as a nontrivial projection operator, then it is strictly speaking the combination of an inclusion operator and the reproducing kernel, since the reproducing kernel is by definition a function in the domain space in both arguments.

[29](Feichtinger and Werther, "Robustness of Regular Sampling in Sobolev Algebras", p. 84), see also the discussion in (Dick and Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Chapter 2) where the reproducing kernel for some special cases of $H^s(\mathbb{R})$ is constructed.

[30]See for example (Boor and Lynch, "On Splines and their Minimum Properties"; Wahba, *Spline Models for Observational Data*; Wahba, *An Introduction to Model Building With Reproducing Kernel Hilbert Spaces*).

Example 2.29. The Paley-Wiener space $\Omega_B(\mathbb{R}^d)$ of B-bandlimited functions in the Fourier domain is a reproducing kernel Hilbert space.[31]

Example 2.30. The space of square-integral, holomorphic functions in the complex plane is a reproducing kernel Hilbert space with the Bergman kernel as reproducing kernel.[32] Note that being holmorphic is closely related to being bandlimited in the Fourier domain, which provides a connection to the previous example.

Example 2.31. For a differential equation that is positive definite and of sufficiently high order, the Green's function is a reproducing kernel. For example, for a self-ajoint elliptic differential equation, the Green's function exists and is a reproducing kernel when the problem is positive definite.[33]

The above examples of reproducing kernel Hilbert spaces include many constructive spaces that are of importance in applications, although this property has not been considered in the literature. We will discuss why this correlation is not accidental in Chapter 4.

In the closing of this section, we will briefly consider the representation of linear operators on reproducing kernel Hilbert spaces where these admit an alternative description.

Proposition 2.37. *Let $\mathcal{H}_k(X)$ be a reproducing kernel Hilbert space defined over a set X with reproducing kernel $k_y(x)$, and let $L : \mathcal{H}_k \to \mathcal{H}_k$ be a linear operator on $\mathcal{H}_k$ with adjoint L^*. Then the action of L for all $f \in \mathcal{H}$ satisfies*

$$(Lf)(y) = \langle f(x), \Lambda(y, x) \rangle$$

*where the **kernel** $\Lambda(x, y)$ **of** L is given by*

$$\Lambda(x, y) = L^* k_y(x).$$

The above proposition is easily established by exploiting linearity. The result shows that in a reproducing kernel Hilbert space any operator can be considered as an "integral operator" for the inner product.

---

[31] Nashed and Walter, "General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces".

[32] The kernel was introduced in (Bergmann, "Über die Entwicklung der harmonischen Funktionen der Ebene und des Raumes nach Orthogonalfunktionen"). It is said that Bergmann discovered it while he was an undergraduate student at a German university, and due to his limited knowledge of German solved an exercise for $\mathbb{C}$ instead of $\mathbb{R}$. For are detailed discussion of the Bergman kernel see for example (Meschkowski, *Hilbertsche Räume mit Kernfunktion*, Chapter 4).

[33] The basic connection between Green's functions and reproducing kernels was already pointed out in Aronszjan's original paper (Aronszajn, "Theory of Reproducing Kernels"), see however also (Aronszajn and Smith, "Characterization of Positive Reproducing Kernels. Applications to Green's Functions.") and (Meschkowski, *Hilbertsche Räume mit Kernfunktion*, Chapter 4).

2.2.4   Computations in Hilbert Spaces[34]

Function spaces, equipped with a norm and an inner product, provide the essential structure to effectively work with functions. The mathematical models encountered in most applications, however, employ infinite dimensional function spaces, while computations on a finite computer require finite dimensional descriptions. And even with a finite function space, we have to obtain a representation that is not continuous but suitable for numerical computations. In this section, we will therefore consider two questions:

1. How to obtain a finite representation of an infinite dimensional signal?

2. How to work numerically with a finite but continuous signal?

The first step will require an approximation, that is a loss of information and error will be introduced, whereas the second step will employ an isomorphism. We will address the steps in order in the following.

Remark 2.29. In the literature, one often refers to the process of going from a possibly infinite dimensional, continuous description to a finite representation on a computer as 'discretization'. However, different and often conflicting interpretations of the term exist, and we will hence avoid this terminology whenever possible.

2.2.4.1   From Infinite to Finite Dimensions

Given a class of infinite dimensional signals, functions arising from a physical process, for example all possible images that can be obtained on a camera sensor, we seek a finite description for them. A first step towards this objective is to obtain an understanding of the characteristics of the functions—to determine the function space the signals lie in. The Lebesgue space L_2 is often an easily justified assumption, and in problems involving partial differential equations often Sobolev spaces arise. However, a "tight" description that precisely characterizes the functions of interest is desirable, since this yields additional structure which allows for stronger results, analogous to the refinement in the description of operators and bases that was possible in the foregoing by adding a norm and an inner product to the "bare" linear spaces we began with. Unfortunately, finding such descriptions is in most cases very challenging, and even for seemingly simple signals such as natural images far from trivial.[35] Henceforth, we will assume a "suitable" infinite dimensional Hilbert space $\mathcal{H}(X)$, and instead return to some aspects of the question in Chapter 4.4.

---

[34]The following presentations is largely our own perspective on the subject. For a more detailed and classical presentations see for example (DeVore and Lorentz, *Constructive approximation*; Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*). A less complete treatment but aimed at light transport simulation can be found in (Cohen and Wallace, *Radiosity and Realistic Image Synthesis*).

[35]See the book by Mallat (*A Wavelet Tour of Signal Processing: The Sparse Way*) for an up-to-date discussion of the subject.

We are faced with the question of how to obtain a finite representation for the elements in our "suitable" function space $\mathcal{H}(X)$—a problem where we have to be satisfied with approximations since we have to reduce the amount of information that is employed, unavoidably progressing from an infinite set of basis function coefficients to a finite dimensional one. The germ to the answer is provided in Def. 2.31. There, we introduced the shorthand notation

$$f = \sum_{i=1}^{n} f_i \, \psi_i \tag{2.40}$$

for the convergence of an infinite series ad infinitum,

$$\left\| f - \sum_{i=1}^{n} f_i \, \psi_i \right\| \xrightarrow[n\to\infty]{} 0. \tag{2.41}$$

But when we accept some error ϵ, which, as we observed before, is unavoidable, then the above series becomes

$$\left\| f - \sum_{i=1}^{n} f_i \, \psi_i \right\| < \epsilon \tag{2.42}$$

and for every $\epsilon > 0$ we can find a finite $n \in \mathbb{N}$ such that the inequality is satisfied, a result we already established in Corollary 2.1.

Remark 2.30. In the foregoing, for example in Eq. 2.41, we always considered **convergence in the norm**, that is the norm between the sequence $\sum_{i=1}^{n} f_i \, \psi_i$ and the function f vanishes at the limit. Unfortunately, this does not characterize the local behaviour at the limit and in general one is not guaranteed that

$$\left| f(x) - \sum_{i=1}^{n} f_i \, \psi_i(x) \right| \xrightarrow[n\to\infty]{} 0 \tag{2.43}$$

for all x in the domain X, even when the sequence $\sum_{i=1}^{n} f_i \, \psi_i(x)$ converges in the norm. When Eq. 2.43 does hold, then one speaks of **pointwise convergence**, and when the rate of convergence is the same for all points x, one speaks of **uniform convergence**. The latter two notions are usually much stronger than convergence in the norm, although, as we have seen in the previous section, for reproducing kernel Hilbert spaces all three are equivalent, testifying for the specialness of these spaces.

From the discussion in the foregoing sections it is easily concluded that suitable sequences $\{\psi_i\}_{i=1}^{k}$ for the approximation in Eq. 2.42 are given by Riesz bases and frames for $\mathcal{H}(X)$, since there the expansion coefficients can be obtained using the dual basis functions. Fixing an arbitrary basis or frame $\{\psi_i\}_{i=1}^{k}$ for $\mathcal{H}(X)$, we obtain a sequence $(\hat{f}_1, \cdots, \hat{f}_m, \cdots, \hat{f}_k)$ of finite dimensional approximation functions

$$\hat{f}_m = \sum_{i \in I_m} f_i \, \psi_i.$$
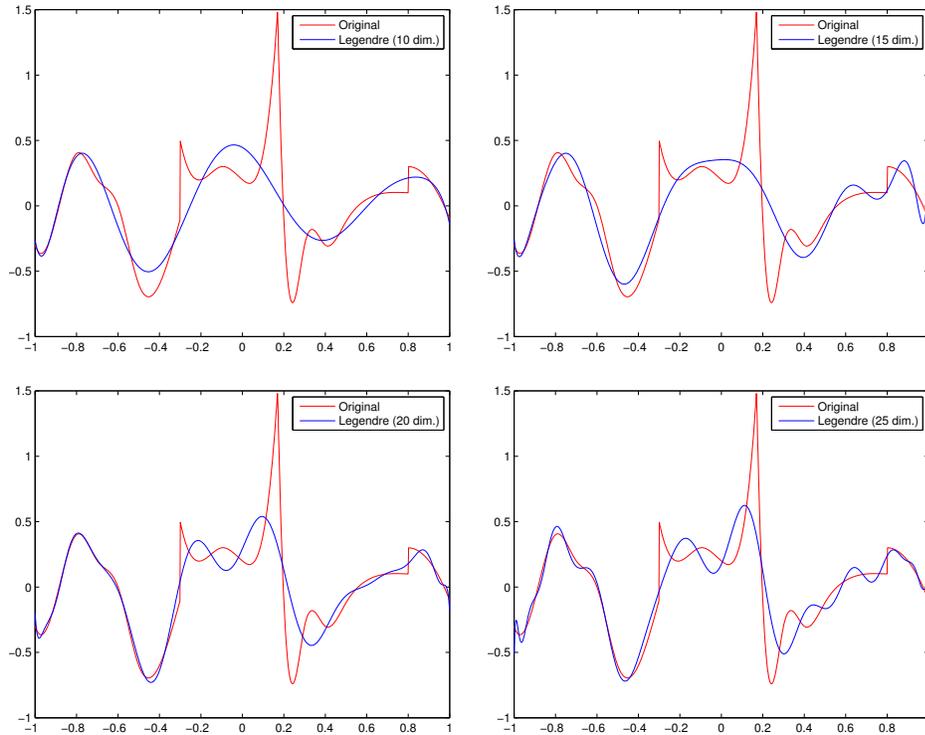
Figure 2.5: Approximation of a signal (red) using Legendre polynomials for approximation spaces with increasing dimensionality.

where I_m is an index set of cardinality m, and we assume $I_m \subset I_{m+1}$, see also Fig. 2.5 for an example. Each of the functions lies in a finite dimensional space $\hat{H}_m$ spanned by the basis function employed in the approximation, and the spaces form a sequence

$$\hat{\mathcal{H}}_1 \subset \ldots \subset \hat{\mathcal{H}}_m \subset \ldots \mathcal{H}.$$

It is important to note that the approximation does depend on the index set chosen, and the m-dimensional index sets are by no means equivalent, although unless $\{\psi_i\}_{i=1}^k$ is an orthonormal basis there is no canonical choice for the set or at least such a choice is not easily determined. We summarize the foregoing discussion in the following definition.[36]

Definition 2.51. *Let $\mathcal{H}(X)$ be a Hilbert space and $f \in \mathcal{H}(X)$ an arbitrary element, and let $\{\psi_i\}_{i=1}^k$ be a Riesz basis or a frame for $\mathcal{H}(X)$, where k is possibly infinity. Then an mth **order approximation** $\hat{f}_m$ of $f \in \mathcal{H}(X)$ for a suitable index set I_m of finite cardinality m is*

$$f \approx \hat{f}_m = \sum_{i \in I_m} f_i\, \psi_i$$

---

[36]See also (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Chapter 9.1).

and $\hat{f}_m$ lies in the **approximation space**

$$\hat{\mathcal{H}}_m(X) = \operatorname*{span}_{i \in I_m} (\psi_i).$$

The **approximation error** $E_m(f)$ *of $\hat{f}_m$ is the residual*

$$E_m(f) = \|f - \hat{f}_m\|^2.$$

For so defined approximations, we have the following result, which follows immediately from Remark 2.10 and Parseval's identity.

Proposition 2.38. *Let $\mathcal{H}(X)$ be an infinite dimensional Hilbert space and $f \in \mathcal{H}(X)$ an arbitrary element, and let $\{\psi_i\}_{i=1}^k$ be a Riesz basis or a frame for $\mathcal{H}(X)$. Then the mth order approximation $\hat{f}_m$ of a function $f \in \mathcal{H}(X)$ is the orthogonal projection of f onto the approximation space $\hat{H}_m$. Moreover, when $\{\psi_i\}_{i=1}^\infty$ is an orthonormal basis then the approximation error is given by*

$$E_m(f) = \sum_{i \notin I_m} f_i.$$

When the signals of interest are defined in an infinite dimensional space, the principal tool to transition from infinite to finite dimensions is hence provided by bases and frames, which justifies the considerable detail we devoted to them in the foregoing. In the literature, one will often find more general notions of an approximation problem, for example for functions in Banach spaces, but the above will suffice for our purposes. In fact, in Chapter 4 we will see that even our current setting, that of general Hilbert spaces, is in a certain sense still too general.

Remark 2.31. The question we leave open in our discussion is which representation are suitable to obtain "good" approximations such that $\|f - \hat{f}_m\|$ is small even when only few basis function coefficients are employed. The problem is studied in approximation theory, and it leads to constructions such as wavelets, curvelets, and ridgelets. For us, however, other considerations will take center stage.[37]

With the finite approximations introduced in Def. 2.51, we have a description of functions that is fundamentally amenable to a treatment on a finite machine. How this can be realized pragmatically will be discussed in the next section.

2.2.4.2   Computations with Continuous Signals

With the results of the previous section, we seek a mapping from a finite but still continuous function space to a representation that is amenable to a computer

---

[37]For the state-of-the-art on the approximation of natural images see the book by Mallat (ibid.). That this setting is still an area of ongoing research shows already that optimal, constructive approximation is challenging, and to our knowledge few conclusive results exist in higher dimensions and for nontrivial domains such as manifolds.

implementation. A suitable mapping is provided by Theorem 2.6 where we showed that a Hilbert space $\mathcal{H}(X)$ is isomorphic to the discrete Lebesgue space ℓ_2^k, and, moreover, that the isomorphism is provided by an orthonormal basis. Thus, for the finite approximation spaces $\hat{\mathcal{H}}_m(X)$ of interest to us, where ℓ_2^m is just Euclidean space $\mathbb{R}^m$, numerical computations can be performed efficiently using linear algebra. Using the analysis and reconstruction operators $\mathcal{A}(\phi)$ and $\mathcal{R}(\phi)$ introduced in Def. 2.18, the situation can be depicted as follows:

<div align="center">

physical   computational

$\mathcal{R}(\phi)$

$(\hat{\mathcal{H}}_m, \langle\,,\rangle) \quad \xleftrightarrow{\hspace{4cm}} \quad (\mathbb{R}^m, \cdot)$

$\mathcal{A}(\phi)$

continuous   linear algebra

</div>

The diagram shows that for an approximation space $\hat{\mathcal{H}}_m(X)$, we can map a problem to Euclidean space $\mathbb{R}^m$, perform numerical computations in this setting using linear algebra, and then reconstruct the result to obtain the sought answer in the continuous domain. We will formalize the mapping of a function to its representation in $\mathbb{R}^m$ as follows.

Definition 2.52. *Let $\mathcal{H}(X)$ be an m-dimensional Hilbert space with orthonormal basis $\{\phi_i\}_{i=1}^n$, and let $f \in \mathcal{H}(X)$ be an arbitrary function in $\mathcal{H}(X)$. Then the* **coefficient vector** $\mathrm{f}(\phi)$ *of f with respect to $\{\phi\}_{i=1}^m$ is*

$$\mathrm{f} = \mathrm{f}(\phi) = (f_1, \cdots, f_n) \in \mathbb{R}^n \cong \ell_2^n$$

where the coefficients f_i are given by $f_i = \langle f, \phi_i \rangle$.

In practice, the coefficient vector f is usually an m-dimensional column vector, and when it arises from a finite approximation $\hat{f}_m$ we will write $\hat{\mathrm{f}}_m$.

Remark 2.32. Most function space of interest for numerical computations are closed subspaces of $L_2(X)$, over some domain X, in which case the basis function coefficients are determined by the integral

$$f_i = \langle f, \phi_i \rangle = \int_X f(x)\,\phi_i(x)\,dx. \tag{2.44}$$

To exemplify how computations are mapped from the continuous to the discrete domain, we will consider a change of basis from the orthonormal basis $\{\phi_i\}_{i=1}^n$ to an arbitrary frame or basis $\{\psi_i\}_{i=1}^m$, an applications which will in fact be of considerable importance in Chapter 4. The change of basis is the projection of $f \in \mathcal{H}(X)$, when represented in $\{\phi_i\}_{i=1}^n$, onto $\{\psi_i\}_{i=1}^m$. Hence, we have

$$f_i(\psi) = \langle f, \psi_i \rangle \tag{2.45a}$$

$$f_i(\psi) = \left\langle \sum_{j=1}^n f_j(\phi)\,\phi_j\,,\,\psi_i \right\rangle \tag{2.45b}$$

and exploiting the linearity of the inner product and that the basis function coefficients are scalars we obtain

$$f_i(\psi) = \sum_{j=1}^{n} f_j(\phi) \langle \phi_j, \psi_i \rangle. \tag{2.45c}$$

The inner product also yields scalars, and hence we can define coefficients $b_{ij} = \langle \phi_j, \psi_i \rangle$ for the projection of the basis function ψ_i onto ϕ_j. Eq. 2.45c can then be written as

$$f_i(\psi) = \sum_{j=1}^{n} b_{ij} \, f_j(\phi) \tag{2.45d}$$

which is just a matrix-vector equation

$$f(\psi) = B(\psi, \phi) \, f(\phi) \tag{2.45e}$$

for the matrix $B(\psi, \phi) = \{b_{ij}\}$. From Eq. 2.45e we also have that the inverse change of basis from $\{\psi_i\}_{i=1}^{n}$ to $\{\phi_i\}_{i=1}^{m}$ is given by the inverse basis matrix, that is

$$f(\phi) = B^{-1}(\psi, \phi) \, f(\psi) = B(\phi, \psi) \, f(\psi). \tag{2.46}$$

whose elements we will denote as b_{ij}^{-1}. When I is the identity matrix, then this is equivalent to

$$B^{-1}(\psi, \phi) \, B(\psi, \phi) = B(\psi, \phi) \, B^{-1}(\psi, \phi) = I, \tag{2.47}$$

stating that the change of variables is reversible, and that the basis function coefficients with respect to $\{\phi_i\}_{i=1}^{n}$ can be "reconstructed" from those with respect to $\{\psi_i\}_{i=1}^{m}$. However, the inverse has an additional property that is important for us. Applying the derivation in Eq. 2.45 in the inverse direction we have

$$f_i(\phi) = \sum_{j=1}^{n} b_{ij}^{-1} \, f_j(\psi) \tag{2.48a}$$

and, with foresight, assuming also the coefficients b_{ij}^{-1} of the inverse basis matrix have the form $b_{ij}^{-1} = \langle \varphi_j, \phi_i \rangle$, for a basis φ_j yet to be determined, yields

$$f_i(\phi) = \sum_{j=1}^{n} \langle \varphi_j, \phi_i \rangle \, f_j(\psi). \tag{2.48b}$$

Exploiting linearity and collecting the terms that are summed over j we have

$$f_i(\phi) = \left\langle \sum_{j=1}^{n} f_j(\psi) \, \varphi_j \, , \, \phi_i \right\rangle. \tag{2.48c}$$

But by definition $f = \sum_{i=1}^{n} f_j(\psi)\,\tilde{\psi}_j$, and for a Riesz basis the φ_j have to be the unique dual basis functions $\tilde{\psi}_j$, while for a frame they have to be some dual frame $\tilde{\psi}_j$. Thus, we obtain

$$f_i(\phi) = \left\langle \sum_{j=1}^{n} f_j(\psi)\,\tilde{\psi}_j\,,\,\phi_i \right\rangle. \tag{2.48d}$$

The derivation shows that the columns of the inverse basis matrix contain the basis function coefficients of the dual basis or frame functions $\{\tilde{\psi}_i\}_{i=1}^{n}$ with respect to the orthonormal "reference" basis $\{\phi_i\}_{i=1}^{m}$. In fact, this could have been concluded more directly, but with less insight, from Eq. 2.47 which in component form reads

$$\sum_{j=1}^{n} b_{ij}\,b_{jk}^{-1} = \delta_{ik} \tag{2.49}$$

which is just a restatement of the biorthogonality condition for the dual basis functions. Importantly, the relationship between the inverse $B^{-1}(\psi,\phi) = B(\phi,\psi)$ of the matrix $B(\psi,\phi)$ and the dual basis functions provides a numerical means to obtain the dual basis or frame functions that is easily applied in practice. We summarize the above considerations in the following definition.

Proposition 2.39. *Let $\mathcal{H}(X)$ be an m-dimensional Hilbert space with orthonormal basis $\{\phi_i\}_{i=1}^{m}$, and let $\{\psi_i\}_{i=1}^{n}$ be an arbitrary basis or frame for $\mathcal{H}(X)$ with $n \geq m$ Then the **basis matrix** $B(\psi,\phi)$ of $\{\psi_i\}_{i=1}^{n}$ with respect to $\{\phi_i\}_{i=1}^{m}$ is*

$$B = B(\psi,\phi) = \{b_{ij}\} = \begin{bmatrix} \langle \psi_1, \phi_1 \rangle & \cdots & \langle \psi_1, \phi_m \rangle \\ \vdots & \ddots & \vdots \\ \langle \psi_n, \phi_1 \rangle & \cdots & \langle \psi_n, \phi_m \rangle \end{bmatrix} \in \ell_2^n \times \ell_2^m \tag{2.50}$$

and it provides the change of basis from $\{\phi_i\}_{i=1}^{m}$ to $\{\psi_i\}_{i=1}^{m}$. The inverse basis matrix

$$B^{-1}(\phi,\psi) = B(\psi,\phi)$$

provides the inverse change of basis from $\{\psi_i\}_{i=1}^{n}$ to $\{\phi_i\}_{i=1}^{m}$, and, moreover,

$$b_{ij}^{-1} = \langle \tilde{\psi}_j, \phi_i \rangle$$

and the columns of the inverse kernel matrix are formed by the basis function coefficients of the dual basis or frame functions $\{\tilde{\psi}_i\}_{i=1}^{n}$.

Let us now briefly comment on how the inverse kernel matrix is determined in the case of a frame.

Remark 2.33. A non-square matrix does not have an inverse in the classical sense but only a left or right pseudo-inverse as introduced in Def. 2.15. For our purposes, this is sufficient since any left inverse will enable us to reconstruct a signal from its frame representation. A suitable choice for the inverse is the Moore-Penrose pseudo-inverse which can be determined using the singular value decomposition and which has minimal L_2 norm.[38] Alternatives exist in the literature,[39] but, unless mentioned otherwise, we will in the following always employ the Moore-Penrose pseudo-inverse.

Remark 2.34. Since basis projection is a *projection*, and it can hence have a nontrivial kernel, care is needed when the identity $B^{-1}(\psi, \phi) B(\psi, \phi) = I$ is interpreted, in particular when it is not clear in which spaces the functions of interest lie.

Remark 2.35. The attentive reader will have noticed that the discussion in this section has an imperfection: we still require continuous operations, such as the computation of integrals, to obtain the discrete representations. For example, when $\mathcal{H}(X) \subset L_2$, then the elements of the basis function coefficients are given by Eq. 2.44, and similarly to obtain the elements of the kernel matrix one has to determine

$$b_{ij} = \langle \phi_j, \psi_i \rangle = \int_X \phi_j(x)\, \psi_i(x)\, dx \qquad (2.51)$$

We will consider this problem in greater detail in Chapter 4, and in fact it provides part of the motivation of the work presented there.

2.2.4.3 Finite Approximations of Operators

In the previous two sections, we considered how to obtain representations of functions that are amenable to numerical computations. However, in many applications the functions are only one part of the problem, and needed is also a numerical description of operators, which for example describe the time evolution of the functions. In the following we will briefly consider this question for Hilbert-Schmidt integral operators, the class of operators we already considered in greater detail in the foregoing.

A Hilbert-Schmidt operator $K : L_2(X) \to L_2(X)$ was in Def. 2.40 defined as

$$g(y) = (Kf)(y) = \int_U f(x)\, k(x, y)\, dx \qquad (2.52)$$

where $k \in L_2(U \times U)$ is a Hilbert-Schmidt kernel. We are interested in a representation of the equation for finite dimensional approximations $\hat{f}_m$ and $\hat{g}_m$

[38](Moore, "On the Reciprocal of the General Algebraic Matrix"; Penrose, "A generalized inverse for matrices"), see (Ben-Israel and Greville, *Generalized Inverses*) for a modern discussion.

[39]For example the Matlab "backslash" operator for an overdetermined system, which is computed by the QR-algorithm, yields an optimally spare pseudo-inverse, see (Moler, *Numerical Computing with MATLAB*, Chapter 5).

in an m-dimensional approximation space $\hat{\mathcal{H}}_m(X)$, and when $\hat{f}_m$ and $\hat{g}_m$ are described by their basis function coefficients with respect to a basis or frame pair $(\psi_i, \tilde{\psi}_i)_{i=1}^{\infty}$ for $L_2(X)$. Representing $\hat{f}_m$ in the basis, Eq. 2.52 becomes

$$g(y) = \int_U \left(\sum_{j \in I_m} f_j \, \psi_j(x) \right) k(x, y) \, dx \qquad (2.53\text{a})$$

where I_m is as before a suitable index set of cardinality m. Exploiting linearity on the right hand side, and since the basis function coefficients f_j are just scalars that do not depend on x, we obtain

$$g(y) = \sum_{j \in I} f_j \int_U \psi_j(x) \, k(x, y) \, dx. \qquad (2.53\text{b})$$

But since $\mathcal{H}_m(X)$ is a closed subspace of $L_2(X)$, and hence the inner product on $\mathcal{H}(X)$ is those on $L_2(X)$, this is equivalent to

$$g(y) = \sum_{j \in I} f_j \, \langle \psi_j(x), k(x, y) \rangle. \qquad (2.53\text{c})$$

Hence, on the right hand side we have the projection of K onto $\mathcal{H}_m(X)$ over its first coordinate, providing a partially finite representation of the operator. To also obtain a finite description in the second coordinate we project the equation onto the basis using the dual basis or frame functions

$$\langle g(y), \tilde{\psi}_i(y) \rangle = \left\langle \sum_{j \in I} f_j \, \langle \psi_j(x), k(x, y) \rangle, \, \tilde{\psi}_i(y) \right\rangle. \qquad (2.53\text{d})$$

Once again exploiting linearity we hence obtain

$$\langle g(y), \tilde{\psi}_i(y) \rangle = \sum_{j \in I} f_j \left\langle \langle \psi_j(x), k(x, y) \rangle, \, \tilde{\psi}_i(y) \right\rangle. \qquad (2.53\text{e})$$

Using Dirac bra-ket notation, this can be written more concisely as

$$\langle g(y), \tilde{\psi}_i(y) \rangle = \sum_{j \in I} f_j \left\langle \psi_j(x) \, \middle| \, k(x, y) \, \middle| \, \tilde{\psi}_i(y) \right\rangle \qquad (2.53\text{f})$$

and by defining the coefficients

$$k_{ij} = \left\langle \psi_j(x) \, \middle| \, k(x, y) \, \middle| \, \tilde{\psi}_i(y) \right\rangle \qquad (2.53\text{g})$$

we obtain the matrix-vector equation

$$g_i = \sum_{j \in I} k_{ij} \, f_j. \qquad (2.53\text{h})$$

We thus derived the following result for the finite representation of a Hilbert-Schmidt operator that can be employed for numerical computations, and which is the analog of Def. 2.51 and Def. 2.52 for functions.

Definition 2.53. *Let $(\psi_i, \tilde{\psi}_i)_{i=1}^k$ be a basis or frame for $L_2(X)$, and let $K :$ $\mathcal{H}(X) \to \mathcal{H}(X)$ be a Hilbert-Schmidt operator with kernel $k(x,y) \in L_2(X \times U)$. The $m^{\textbf{th}}$ **order operator matrix** $\hat{\mathrm{K}}_m : \hat{\mathcal{H}}_m \to \hat{\mathcal{H}}_m$ **of** K is*

$$
\hat{\mathrm{K}}_m = \begin{bmatrix} \left\langle \psi_1(x) \,\middle|\, k(x,y) \,\middle|\, \tilde{\psi}_1(y) \right\rangle & \cdots & \left\langle \psi_1(x) \,\middle|\, k(x,y) \,\middle|\, \tilde{\psi}_m(y) \right\rangle \\ \vdots & \ddots & \vdots \\ \left\langle \psi_m(x) \,\middle|\, k(x,y) \,\middle|\, \tilde{\psi}_1(y) \right\rangle & \cdots & \left\langle \psi_m(x) \,\middle|\, k(x,y) \,\middle|\, \tilde{\psi}_m(y) \right\rangle \end{bmatrix}
$$

where the matrix elements of $\hat{\mathrm{K}}_m$ are defined by

$$
k_{ij} = \left\langle \psi_j(x) \,\middle|\, k(x,y) \,\middle|\, \tilde{\psi}_i(y) \right\rangle = \left\langle \langle \psi_j(x), k(x,y) \rangle \,,\, \tilde{\psi}_i(y) \right\rangle
$$

and $i, j \in I_m$ for a suitable index set I_m of cardinality m. $\hat{\mathrm{K}}_m$ provides the action of K on the approximation space $\hat{\mathcal{H}}_m$ induced by the index set I_m.

Remark 2.36. We can write the explicit derivation of the operator matrix in Eq. 2.53 also more concisely using operator notation. A Hilbert-Schmidt operator is then given by

$$
g = K f. \tag{2.54a}
$$

and when $\hat{\mathrm{f}}_m$ is the coefficient vector for the m-term approximation $\hat{f}_m$ of f, then the above equation becomes

$$
g = K \left(\mathcal{R}_m \, \hat{\mathrm{f}}_m \right) \tag{2.54b}
$$

where $\mathcal{R}_m = \mathcal{R}_m(\psi_I)$ is the reconstruction operator for $\{\psi_i\}_{i \in I}$. The basis projection is obtained using the analysis operator $\mathcal{A}_m = \mathcal{A}_m(\psi_I)$ for the m-term approximation, and hence

$$
\mathcal{A}_m \, g = \mathcal{A}_m \left(K \left(\mathcal{R}_m \, \hat{\mathrm{f}}_m \right) \right). \tag{2.54c}
$$

Rearranging brackets we obtain

$$
\hat{\mathrm{g}}_m = \left(\mathcal{A}_m \, K \, \mathcal{R}_m \right) \hat{\mathrm{f}}_m \tag{2.54d}
$$

which is equivalent to Eq. 2.53h. The above derivation shows that our ansatz is a form of Galerkin projection, which is a very general approach to obtain finite representations for operators, and which for example includes also finite element methods.[40]

Remark 2.37. It should be noted how in the above derivations the linearity of the operator and the linearity of the basis representation were employed hand in hand to obtain the result, and from a computational point of view much of the convenience of linear operators lies in this fact.

---

[40]Cf. (Arnold, Falk, and Winther, "Finite element exterior calculus: from Hodge theory to numerical stability").
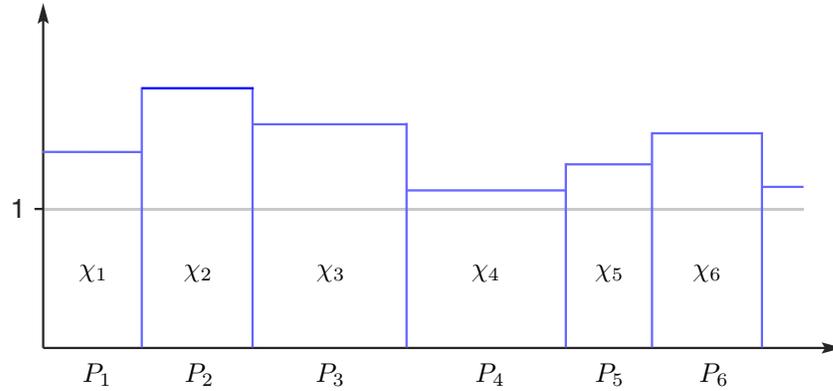
Figure 2.6: Signal represented in a characteristic basis $\{\chi_i\}_{i=1}^k$ for a nonuniform and irregular partition $\{P_i\}_{i=1}^k$.

2.2.5 Examples

In this section, we will consider some examples of Hilbert spaces, which will also play an important role in the following.

2.2.5.1 Characteristic Basis

The characteristic basis is arguably the simplest possible basis, see Fig. 2.6. It is defined over a partition of the domain X, and we will hence begin by formalizing this notion.

Definition 2.54. *Let X be a set, and (X, Σ, μ) a suitable measure space over X.[41] A set $P = \{P_i\}$ formed by elements P_i is a **partition of** X when*

i) $\bigcup P_i = X$;

ii) $P_i \cap P_j = \varnothing$, $i \neq j$.

*A partition is **regular** when $T(P_i) = P_j$, where T is the natural translation operator for X, if it exists, and it is **uniform** when $\mu(P_i) = \mu(P_j)$.*

A set X will in general admit an infinite number of partitions, and although these are mathematically equivalent, they can yield significant differences in practice. As should be clear from the definition, the existence of regular partitions depends on the domain and a well defined translation operator.

---

[41]We recall that a measure space (X, Σ, μ) is a set X together with a σ-algebra defined over X and a possibly weighted measure, see for example (Rudin, *Principles of Mathematical Analysis*, Chapter 11) for details. Readers not familiar with measure theory should in the following assume that $\mu(X)$ is defined as $\mu(X) = \int_X dx$ and it determines the area or volume of X, and, as usual, it is instructive to keep the example $X = [a, b] \subset \mathbb{R}$ in mind.

Definition 2.55. *Let X be a set and (X, Σ, μ) a suitable measure space over X, and let $P = \{P_i\}$ be a partition of X of cardinality n. Then the* **characteristic basis** Ξ_n *induced by P is*

$$\Xi_n = \Xi_n(P) = \{\chi_i\}_{i=1}^{n}$$

where $\chi_i : X \to \mathbb{R}$ is the **characteristic function** *or* **indicator function** *for the set P_i given by*

$$\chi_i(x) = \chi(P_i)(x) = \begin{cases} 1 & x \in P_i \\ 0 & x \notin P_i \end{cases}.$$

The normalized characteristic basis $\bar{\Xi}_n = \{\bar{\chi}_i\}_{i=1}^{n}$ is spanned by the normalized characteristic basis functions

$$\bar{\chi}_i(x) = \frac{1}{\sqrt{\mu(P_i)}} \chi_i(x).$$

With slight abuse of notation, we will in the following often identify the basis with the space it spans, and it is important to keep in mind that a characteristic basis is specific to a partition. An important result for the characteristic basis is the following.

Theorem 2.10. *Let X be a set and (X, Σ, μ) a suitable measure space over X, and let $\mathcal{P} = (P^1, P^2, \ldots, P^n, \ldots)$ be an infinite sequence of partitions of X where the nth element has cardinality n. Then the characteristic basis $\Xi_n(P^n)$ is dense in $L_2(X, \Sigma, \mu)$ as $n \to \infty$, and the Ξ_n are closed subspaces of $L_2(X, \Sigma, \mu)$.*

The theorem is a corollary of the well known result that the Haar basis is dense in L_2.[42]

Corollary 2.8. *Let X be a set and (X, Σ, μ) a suitable measure space over X, and let $P = \{P_i\}$ be a partition of X of cardinality n. Then the space spanned by the characteristic basis $\Xi_n(P) = \{\chi_i\}_{i=1}^{n}$ defined by P is a Hilbert space with inner product*

$$\langle f, g \rangle = \int_X f \, g \, dx$$

It is apparent from the definition that the normalized characteristic basis forms an orthonormal basis for Ξ_n.

Remark 2.38. What is often referred to as "discrete" in the literature, and how the terminology was originally used, is when signals or operators are represented using the characteristic basis. As we mentioned before, the usage is far from uniform in different communities, and we will hence avoid it whenever possible.

---

[42](Haar, "Zur Theorie der Orthogonalen Funktionensysteme"); and we exclude "exotic" sets X where the result might not hold.

By Proposition 2.36, the reproducing kernel $k_y(x) = k(y, x)$ for the space Ξ_n spanned by characteristic functions is

$$k(y, x) = \sum_{i=1}^{n} \bar{\chi}_i(y)\,\bar{\chi}_i(x) = \frac{1}{\mu(P_i)} \sum_{i=1}^{n} \chi_i(y)\,\chi_i(x) = \left\{ \begin{array}{ll} 1 & x, y \in P_i \\ 0 & \text{otherwise} \end{array} \right. \quad (2.55)$$

and the kernel is given by the basis function χ_i for $y \in P_i$. Conversely, by Eq. 2.55 we can think of a characteristic basis as a basis formed by reproducing kernel functions, an interpretation which will be of central importance in Chapter 4.

Remark 2.39. On first sight, it might appear surprising that the point evaluation functional can be continuous on a space formed by discontinuous functions. However, the point evaluation functional is continuous as an object acting on functions in the space, for a fixed partition, and by keeping this in mind the continuity is easily understood.

2.2.5.2 Spherical Harmonics[43]

Spherical harmonics are the eigenfunctions of the Laplace-Beltrami operator $\Delta(S^2)$ on the sphere, and hence the analogues of the Fourier basis functions on S^2.[44] The eigenvalues $\lambda_l = l(l+1)$ of $\Delta(S^2)$ have multiplicity $2l+1$ and orthonormal bases for the $(2l+1)$-dimensional eigenspaces $\mathcal{H}_l$ are provided by Legendre spherical harmonics

$$y_{lm}(\theta, \phi) = \eta_{lm}\, P_{lm}(\cos\theta) \left\{ \begin{array}{ll} \sin(|m|\,\phi) & m < 0 \\ 1 & m = 0 \\ \cos(m\,\phi) & m > 0 \end{array} \right. \quad (2.56)$$

where in each band $-l \le m \le l$.[45] The $P_{lm}(t)$ in Eq. 2.56 are associated Legendre polynomials, and η_{lm} is a normalization constant which is chosen such that the y_{lm} are orthonormal. The union of all bands $\mathcal{H}_l$ is the Hilbert space

$$L_2(S^2) = \bigoplus_{l=0}^{\infty} \mathcal{H}_l = \left\{ f : S^2 \to \mathbb{R} \mid \|f\| < \infty \right\} \quad (2.57)$$

with the usual L_2 inner product

$$\langle f, g \rangle = \int_{S^2} f(\omega)\, g(omega)\, d\omega \quad (2.58)$$

---

[43]A more detailed discussion of the theory of spherical harmonics and proofs for our claims can for example be found in the excellent books by Freeden and co-workers (Freeden, Gervens, and Schreiner, *Constructive Approximation on the Sphere (With Applications to Geomathematics)*; Freeden and Schreiner, *Spherical Functions of Mathematical Geosciences*).

[44]The Laplace-Beltrami operator will be formally introduced in Def. 2.152.

[45]In the literature in applied fields, Legendre spherical harmonics are often simply denoted as spherical harmonics.

over the sphere. An orthonormal basis for $L_2(S^2)$ is thus given by all spherical harmonics y_{lm} with $l = 1 \ldots \infty$ and $-l \le m \le l$.

The well known translational invariance of the Fourier basis over the real line corresponds to the rotational invariance of the spaces $\mathcal{H}_l$ on the sphere: $f_l \in \mathcal{H}_l$ implies $\bar{f}_l = \mathrm{R} f_l \in \mathcal{H}_l$ where $\mathrm{R} \in \mathrm{SO}(3)$ is an arbitrary rotation and the action of the rotation group $\mathrm{SO}(3)$ is defined pointwise as $\mathrm{R} f = f \circ \mathrm{R}^{-1}$.

An important result in the theory of spherical harmonics is the addition theorem

$$P_l(\cos \gamma) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} y_{lm}(\bar{\omega}) \, y_{lm}(\omega), \qquad (2.59)$$

where $P_l(t) = P_{l0}(t)$ is the Legendre polynomial of degree l and $\gamma = \angle(\bar{\omega}, \omega)$. A comparison to Proposition 2.36 shows that Eq. 2.59 provides an expansion of the reproducing kernel for $\mathcal{H}_l$ in Legendre spherical harmonics,

$$k_l(\bar{\omega}, \omega) = \sum_{m=-l}^{l} y_{lm}(\bar{\omega}) \, y_{lm}(\omega) = \frac{2l+1}{4\pi} P_l(\bar{\omega} \cdot \omega) \qquad (2.60)$$

and the kernel is known as zonal harmonic since it is centered around an axis $\bar{\omega} \in S^2$.

2.2.5.3 Spherical Slepian Functions[46]

An orthogonal basis that spans the same spaces $\mathcal{H}_{\le L} = \bigoplus_{l=1}^{L} H_l$ spanned by finite spherical harmonics expansions, but that has additional properties often useful in applications are spherical Slepian functions. The basis provides a constructive solution to the spatio-spectral concentration problem on the sphere that asks for the optimal space localization of a signal bandlimited in the spherical harmonics domain. Much of the theory was developed for Fourier bandlimited functions over the real line in a series of classic papers by Slepian, Pollak, and Landau,[47] and the results were recently extended to the sphere for functions bandlimited in the spherical harmonics domain by Simons and co-workers.[48]

---

[46]Our presentation follows the exposition in (Simons, Dahlen, and Wieczorek, "Spatiospectral Concentration on a Sphere"). The figures in this section were generated using an extension of the implementation of spherical Slepian functions by Frederik Simons available at http://www.dgp.toronto.edu/people/lessig/effective-dimension/.

[47](Slepian and Pollak, "Prolate Spheroidal Wavefunctions Fourier Analysis and Uncertainty I"; Landau and Pollak, "Prolate Spheroidal Wavefunctions, Fourier Analysis and Uncertainty II"; Landau and Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty III"; Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty V: The Discrete Case"), an introduction to the early work on the subject can be found in two articles by Slepian (Slepian, "On Bandwidth"; Slepian, "Some Comments on Fourier Analysis, Uncertainty and Modeling").

[48]Simons, Dahlen, and Wieczorek, "Spatiospectral Concentration on a Sphere"; Simons, Hawthorne, and Beggan, "Efficient analysis and representation of geophysical processes using localized spherical basis functions"; Simons, "Slepian Functions and Their Use in Signal Estimation and Spectral Analysis".

Spatio-Spectral Concentration on the Sphere Many applications require the efficient representation of signals defined over a subset $U \subset S^2$ of the sphere. For example, in geo- and climate science functions are often restricted to continents or oceans, and for light transport many signals are naturally defined for the hemisphere. Compact representations for a subset $U \subset S^2$ are obtained with functions localized in space, while bandlimited expansions provide the advantages of spherical harmonics such as rotational invariance. However, the analogue of the Fourier uncertainty principle for the sphere shows that both properties are incompatible and no bandlimited representation can be localized in space.[49] To retain the advantages of spherical harmonics while being able to efficiently represent spatially localized signals, we seek bandlimited functions

$$g \equiv g_L = \sum_{l,m}^{L} g_{lm}\, y_{lm} \in \mathcal{H}_{\leq L} \tag{2.61}$$

which maximize the concentration measure

$$\lambda = \frac{\|g\|_U^2}{\|g\|_{S^2}^2} = \frac{\int_U |g|^2 d\omega}{\int_{S^2} |g|^2 d\omega} \tag{2.62}$$

for arbitrary but fixed regions $U \in S^2$. Expanding Eq. 2.62 in spherical harmonics yields

$$\lambda = \frac{\sum_{lm} \sum_{l'm'} g_{lm}\, g_{l'm'}\, d_{lm\, l'm'}}{\sum_{lm} g_{lm}} \tag{2.63}$$

where we defined

$$d_{lm\, l'm'} = \int_U y_{lm}(\omega)\, y_{l'm'}(\omega)\, d\omega.$$

Eq. 2.63 is the spatio-spectral concentration problem on the sphere in the frequency domain. By re-arranging the basis function coefficients in vector form g, and the local correlation coefficients $d_{lm,l'm'}$ as a matrix D, it can be stated more concisely as

$$\lambda = \frac{\mathrm{g}^T \mathrm{D}\, \mathrm{g}}{\mathrm{g}^T \mathrm{g}}. \tag{2.64}$$

Eq. 2.64 is a matrix variational problem and it is known[50] that vectors g_i which render the problem stationary satisfy the eigenvalue equation

$$\mathrm{D}\, \mathrm{g}_i = \lambda_i\, \mathrm{g}_i. \tag{2.65}$$

---

[49](Freeden, Gervens, and Schreiner, *Constructive Approximation on the Sphere (With Applications to Geomathematics)*, Theorem 5.5.1); for the original statement for the real line see for example (Daubechies, *Ten Lectures on Wavelets*; Mallat, *A Wavelet Tour of Signal Processing*).

[50]Horn and Johnson, *Matrix Analysis*.
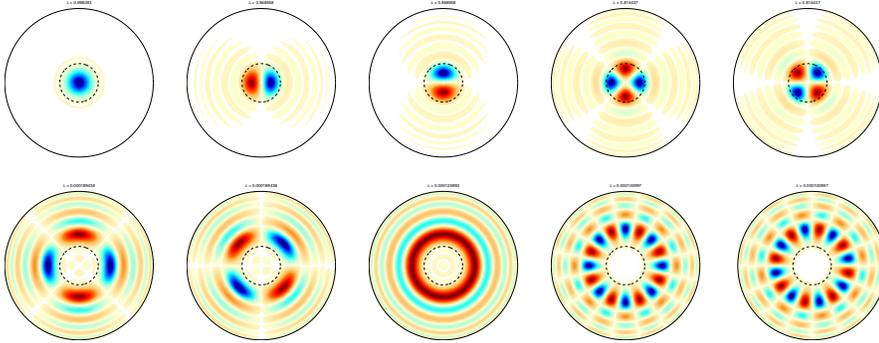
Figure 2.7: Spherical Slepian functions for $L = 20$ and a spherical cap with $\theta \leq 15°$ (dotted circle). The first row shows the first five Slepian functions corresponding to the five largest eigenvalues, and the second row Slepian functions for eigenvalues very close to zero. Positive values are shown in blue and negative ones in red.

In the spatial domain, Eq. 2.65 becomes

$$\int_U D(\bar{\omega}, \omega)\, g_i(\omega) d\omega = \lambda_i\, g_i(\bar{\omega}) \tag{2.66}$$

where the kernel $D(\bar{\omega}, \omega)$ is given by

$$D(\bar{\omega}, \omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} y_{lm}(\bar{\omega})\, y_{lm}(\omega) = \sum_{l=0}^{L} \frac{2l+1}{4\pi} P_l(\bar{\omega} \cdot \omega)$$

which is the reproducing kernel for $\mathcal{H}_{\leq L}$ but with the domain of integration restricted to U, cf. Eq. 2.59.

The functions $g_i(\omega) \in \mathcal{H}_{\leq L}$ in Eq. 2.66 are known as spherical Slepian functions and it follows from the symmetry and positivity of $D(\bar{\omega}, \omega)$ that these can be chosen to be orthogonal over U, cf. Remark 2.14. Additionally, the $\{g_i\}_{i=1}^{n}$ form an orthonormal basis for $\mathcal{H}_{\leq L}$ with $n = (L+1)^2$. Hence, the functions satisfy two orthogonality conditions

$$\int_{S^2} g_i\, g_j\, d\omega = \delta_{ij} \qquad \int_U g_i\, g_j\, d\omega = \lambda_i\, \delta_{ij}.$$

By Eq. 2.62, the eigenvalues λ_i provides a measure for the spatial concentration of the Slepian functions g_i, and by construction g_j is the maximally concentrated function in $\mathcal{H}_{\leq L}$ which is orthogonal to all g_i with $i < j$. Moreover, it follows from the Fourier uncertainty principle and the properties of $D(\bar{\omega}, \omega)$ that the eigenvalues satisfy $1 > \lambda_i \geq \cdots \geq \lambda_n > 0$.

The spectrum of the spatio-spectral concentration problem has a characteristic shape: The first N eigenvalues are close to unity, followed by a region of exponential decay from one to zero, and the remaining eigenvalues are negligible, cf. Fig. 2.8. For the real line it has been shown that for a bandlimit W and a
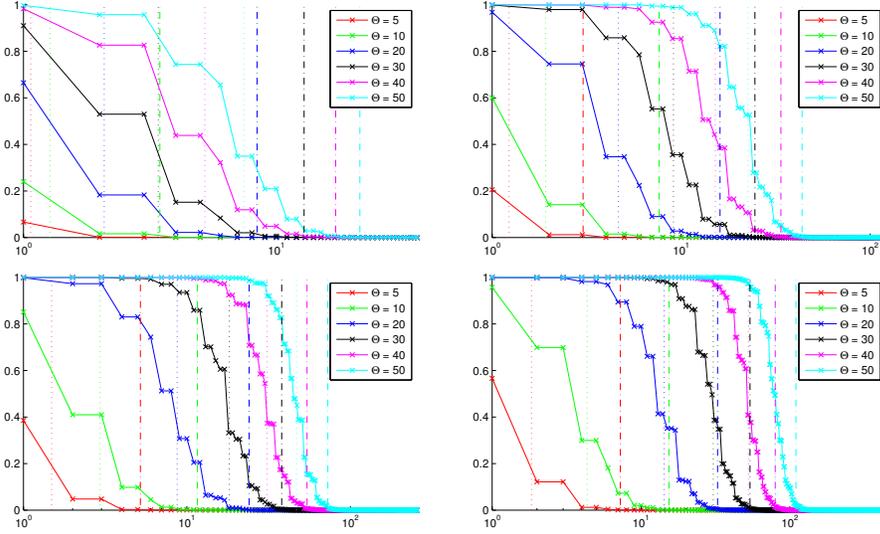
Figure 2.8: Spectrum of the spatio-spectral concentration problem on the sphere for spherical caps with $\theta \leq \Theta$ and $L = 5, 10, 15, 20$ (left to right, top to bottom). The eigenvalue index is shown on the X axis and the magnitude $|\lambda_i|$ of the eigenvalues on the Y axis. Shown are also the Shannon number N (dotted) and generalized Shannon number N_g (dash-dotted), the latter one computed with $B(\partial U) = \lg\left((L+1)^2 |\partial U|^2\right) / \lg(2\pi)$. Clearly visible is the importance of the second term in N_g when the region of spatio-spectral concentration is small.

region of concentration T, the number of eigenvalues greater than $\epsilon \approx 0$ satisfies

$$N(\epsilon) = \underbrace{\frac{W\,T}{2\pi}c}_{\lambda_i \approx 1} + \underbrace{\left(\log\left(\frac{1-\epsilon}{\epsilon}\right)\frac{\mu\,\nu}{\pi^2}\right)\log c}_{1 > \lambda_i > 0} + \underbrace{o(\log c)}_{\lambda_i \approx 0} \qquad (2.67)$$

where c is a scaling parameter for the spatio-spectral region of concentration, and the three terms in Eq. 2.67 correspond to the three characteristic parts of the spectrum.[51] In the second summand, μ and ν determine the numbers of intervals in the frequency and time domain which together have size W and T, respectively, and they measure the size of the boundary of the region of concentration. The first term in Eq. 2.67 is known as the Shannon number N, and we will refer to the first two terms as generalized Shannon number N_g. Unfortunately, for the sphere no results analogous to Eq. 2.67 have been established to date and the rather technical nature of the proof for the real line prevents a straight forward extension to S^2.[52] However, the analogy to

---

[51] Slepian, "Some Asymptotic Expansions for Prolate Spheroidal Wave Functions"; Landau, "The Eigenvalue Behavior of Certain Convolution Equations"; Landau and Widom, "Eigenvalue Distribution of Time and Frequency Limiting".

[52] Landau and Widom, "Eigenvalue Distribution of Time and Frequency Limiting".

this domain and a two-dimensional setting studied by Slepian[53] as well as experimental results[54] suggests that

$$N(\epsilon) = \frac{C}{4\pi} + \log\left(\frac{1-\epsilon}{\epsilon}\right) B(\partial U) \log(C) + o(\log C) \qquad (2.68)$$

where $B(\partial U)$ is a function which depends on the boundary ∂U of the region of concentration and $C = (L+1)^2 A(U)$, cf. Fig. 2.8. The first summand of Eq. 2.68 is again known as the (spherical) Shannon number and it is the dominant term of $N(\epsilon)$ for a sufficiently large region of spatio-spectral concentration. The spectrum is then well described by a step function with the discontinuity at N, and a function localized in U can be represented accurately with only the first N Slepian functions which are all well confined in the region of concentration, cf. Fig. 2.7. If the spatio-spectral region of concentration is small, then a non-negligible contribution to the total spectrum is contained in the transition region from one to zero. In this case, the number of Slepian functions which is needed to accurately represent a localized signal is given by the generalized Shannon number N_g, and because the Slepian basis functions are no longer well concentrated the signal representation will have significant leakage into the complement $S^2 \setminus U$ of the region of concentration U.

The problem dual to optimally concentrating a bandlimited function in a region $U \subset S^2$ is optimally concentrating a signal in the frequency domain which is localized in $U \subset S^2$. The solution to this problem is again provided by spherical Slepian functions.

Computation of Slepian Functions   The space $\mathcal{H}_{\leq L}$ of bandlimited functions on the sphere is finite dimensional. Spherical Slepian functions can thus be obtained by solving Eq. 2.65 using standard numerical algorithms, although in practice some care is required since the eigenvalues are clustered and the computation of the eigenvectors is hence numerically delicate. In the case of a polar spherical cap, the computations are simplified by the existence of a tri-diagonal differential operator that commutes with $D(\omega, \bar{\omega})$, and hence has the same eigenvectors, and which has a well behaved spectrum.[55]

---

[53]Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty IV: Extensions to many Dimensions; Generalized Prolate Spheroidal Functions".

[54]Simons, Dahlen, and Wieczorek, "Spatiospectral Concentration on a Sphere"; Simons, *Personal Communication*.

[55]See (Grünbaum, Longhi, and Perlstadt, "Differential Operators Commuting with Finite Convolution Integral Operators: Some Non-Abelian Examples") and also the discussion in (Simons, Dahlen, and Wieczorek, "Spatiospectral Concentration on a Sphere").

2.3 Geometric Mechanics[56]

Geometric mechanics employs the mathematical tools developed for quantum mechanics and the general theory of relativity, Lie groups and nonlinear analysis on manifolds, to gain insight into the structure of mechanical systems, and the theory elucidates and unifies many aspects of classical and modern physics—with a common mathematical description for systems such as quantum mechanics, quantum field theory, fluid dynamics, rigid body motion, galaxy formation, and general relativity—leading to a renewed interest in the "old physics" in the past 40 years. Moreover, recent efforts shows that the structure revealed by geometric mechanics is critical for the development of numerical algorithms that discretely preserve the essential aspects of a continuous system, and hence enable simulations that are qualitatively correct irrespective of simulation parameters or implementation details.

Geometric mechanics describes the dynamics of a system as curves on a manifold. Depending on the "dialect" used, Lagrangian mechancis or Hamiltonian mechanics, this is either the configuration or the phase space of the system. When combined with Lie groups and their action, the geometric structure enables to describe and analyze symmetries—invariances under the group action—and naturally leads to conserved quantities and reduced mechanical systems described by less variables. For so called Lie-Poisson systems, where the configuration space is a Lie group, the dynamics can be completely reduced to the Lie algebra, providing a particularly elegant and concise description. Classical examples of Lie-Poisson systems are the Euler top and the ideal Euler fluid, and we will show that light transport, in an idealized setting, also belongs to this distinguished class.

The mathematical fundament which is required to formulate light transport as a Hamiltonian system and to study its symmetries, which we will undertake in Chapter 3, is presented in this section. Background on nonlinear tensor analysis on manifolds is introduced in Chapter 2.3.2, while in Chapter 2.3.4 the modern formulation of Hamiltonian mechanics is discussed, and reduction theory and the description of mechanical systems with symmetry is introduced in Sec. 2.3.5. Before we get submersed in technical details, however, we will briefly discuss the history of geometric mechanics, which will help to understand its current position in the sciences, and in Sec. 2.3.1 we will outline the conceptual ideas of the formulation, which will provide further motivation and context for the technical language introduced afterwards.

A Short History of Geometric Mechanics Geometric mechanics is concerned with the symmetries and geometric structures that underlie physical

---

[56]The material in this section is principally drawn from the writings of Jerry Marsden, Tudor Ratiu, and co-workers (*Foundations of Mechanics*; *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*; *Manifolds, Tensor Analysis, and Applications*; *Mechanical Systems: Symmetry and Reduction*), with some additions from the books by Frankel (*The Geometry of Physics*). References are largely omitted in this section although no claim of originality of the presented material is made whatsoever.

systems. The roots of this interpretation of mechanics can be traced back to the work of Leibniz and Newton, who already exploited symmetries and conserved quantities for the solution of the two body problem,[57] and much of the modern theory was anticipated in the work by early mechanicists such as Fermat, Pascal, Euler, Lagrange, Laplace, Gauss, Cauchy, Legendre, Jacobi, and Hamilton.[58] Lagrange, for example, had more than 40 years before Hamilton already a sophisticated picture of Hamiltonian mechanics, and he was for instance aware of the conservation of the symplectic 2-form on extended configuration space.[59] Ironically, Hamilton's most important contribution to mechanics was the principle of least action in Lagrangian mechanics, although his work on indeed laid the foundation for Hamiltonian mechanics. Contemporary to Hamilton, Jacobi developed yet an alternative description of mechanics on phase space, which is known today as Hamilton-Jacobi theory.

In the middle of the 18th century, for example in the thinking of such preeminent scientists as Helmholtz and Maxwell, a shift in the understanding of the nature of the natural sciences developed.[60] While scientists were always concerned with the description of the true working principles of the physical world, one then began to consider also an analogy, a formal illustration, as valuable—if it could provide new insights and enabled verifiable predictions.[61] This shift in perspective led Maxwell to his famous equations, and it opened up the horizon for a more abstract and conceptual description of mechanics. Hertz's understanding of mechanics reflected this new thinking, as is evident in his re-development of mechanics using a principle of least curvature instead of Newton's force concept,[62] and he was probably also the first to speak of a "Geometrisierung der Mechanik". Also in the second half of the 18th century, chiefly with the work of Riemann and Klein, the concept of non-Euclidean geometry was developed and the connection between geometry and symmetry became apparent. Lie developed this into a coherent theory of continuous symmetries, and his work contains already the essentials of what is today reduction theory, although he was apparently unaware on its significance for the description of mechanical systems. The classical period of mechanics culminated in Poincaré's work who's understanding of mechanics reached a remarkable proximity to the modern, geometric formulation. This is exemplified by his description of the dynamics of a rigid body and a fluid where he employs in both cases what is now known as Euler-Poincaré equation, and in a form almost identical to modern treatments,[63] a formulation which was developed further in

---

[57]See (Singer, *Symmetry in Mechanics: A Gentle, Modern Introduction*) for a modern account of Newton's applications of symmetries and conserved quantities for the solution of the two body problem.

[58]Cf. (Synge, "Focal Properties of Optical and Electromagnetic Systems", Section 1).

[59]See (Marsden, "Remarks on Geometric Mechanics") for a more detailed discussion.

[60]See (Klein, *Paul Ehrenfest*).

[61]Exemplary for this new approach was Maxwell's vortex model of "atoms".

[62]Hertz, *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*.

[63](Poincaré, "Sur une forme nouvelle des équations de la méchanique"), and compare for example p. 10 in (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*) to p. 371 in Poincaré's paper.

the dissertations of Ehrenfest and Hamel.

In the first decades of the 20th century, the long and fruitful development of classical mechanics came to a sudden halt when the "new theories"—Einstein's theory of relativity and quantum mechanics—captured the attention of physicists and mathematician.[64] This only changed in the 1960s when work on quantization, trying to bridge the gap between classical and quantum mechanics, required to also revisit the "old theory".[65] This led to a "renaissance of classical mechanics"[66] when it was realized, among others by Kostant, Souriau, Kirillov, Arnold, Abraham, Guillemin, Sternberg, and Smale, that the mathematics and concepts developed and popularized for the modern physical theories, their geometric and (Lie) group theoretic interpretation, provided a language to reformulate mechanics into a more insightful and consistent mathematical theory.[67] With hindsight, this reformulation provided the mathematization of mechanics Hilbert had sought 60 years earlier,[68] and the insight furnished by the modern perspective enabled to answer some long standing questions, such as a canonical formulation of the rigid body and ideal fluid dynamics. The work was continued for example by Arnold, Marsden, Weinstein, and Ratiu in whose hands geometric mechanics developed into a well established subject, and who generalized the theory to include a remarkable range of physical systems from classical ones such as the rigid body,[69] the heavy top,[70] falling cats,[71], ideal fluid dynamics,[72], and elasticity[73] to moderns ones such as magnetohydrodynamics,[74]

---

[64]The rare exceptions are for example Hamel's work (*Theoretische Mechanik*) and those by Synge ("On the Geometry of Dynamics").

[65]See for example (van Hove, "Sur le probléme des relations entre les transformations unitaires de la Mécanique quantique et les transformations canoniques de la Mécanique classique."; van Hove, *On Certain Unitary Representations of an Infinite Group of Transformations*) and the early work by Kostant, Souriau, Guillemin, Sternberg, and Kirillov.

[66]Butterfield, "On Symplectic Reduction in Classical Mechanics", p. 2.

[67]The classic references are the books by Abraham and Marsden (*Foundations of Mechanics*) and Arnold (*Mathematical Methods of Classical Mechanics*), whose first editions appeared in the 1970s. An interesting early text on these developments from the point of view of the physics is those by Sudarshan and Mukunda (*Classical Dynamics: A Modern Perspective*). An outstanding contribution to the mathematics employed for geometric mechanics was made by Cartan, cf. (Cartan, *Differential Forms*).

[68]The axiomatization of mechanics was part of Hilbert's sixth problem (Hilbert, "Mathematical Problems").

[69]See for example (Arnold, "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits").

[70]See for example (Cushman and Bates, *Global Aspects of Classical Integrable Systems*, Chapter V).

[71]Marsden, *Lectures on Mechanics*; Marsden and Ostrowski, "Symmetries in Motion: Geometric Foundations of Motion Control".

[72]See for example (Arnold, "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits"; Ebin and Marsden, "Groups of Diffeomorphisms and the Motion of an Incompressible Fluid"; Arnold and Khesin, *Topological Methods in Hydrodynamics*).

[73]See for example (Marsden and Hughes, *Mathematical Foundations of Elasticity*; Simo, Marsden, and Krishnaprasad, "The Hamiltonian structure of nonlinear elasticity: The material and convective representations of solids, rods, and plates"; Kanso et al., "On the geometric character of stress in continuum mechanics").

[74]Marsden and Morrison, "Noncanonical Hamiltonian Field Theory and Reduced MHD".

plasma physics,[75] and general relativity.[76] This flourished a new self-conception of classical mechanics as the mathematical theory of systems that are described by a generalized Lagrangian and Hamiltonian formalism:[77]

> *"Even more striking are true statements like this: Don't tell me that quantum mechanics is right and classical mechanics is wrong—after all, quantum mechanics is a* special case *of classical mechanics."*

Geometric mechanics differs from classical descriptions such as Newtonian mechanics not only by the mathematical language that is employed, but also in the quest for understanding and insight into the fabric underlying the physics: calculations are no longer sufficient, but one wants to understand the structure which explains why a calculation provides the sought answer, in fact one wants to derive the equations from the structure,[78] or using the words of a philosopher:[79]

> *"[...] classical mechanics is alive and kicking, not least through deepening our understanding of time-honoured systems such as the rigid body— whose analysis in traditional textbooks can be all too confusing!"*

In geometric mechanics some effort is needed for the analysis of even the simplest systems, such as a particle, while Newtonian mechanics provides readily applicable answers in these cases. However, the modern perspective explains *why* the equations have their form, and this enables to employ the ansatz used for a particle for a wide range of systems, from quantum mechanics to general relativity.

Surprisingly, the geometric viewpoint is also of central importance for the numerical simulations of mechanical systems. One of the first applications which made this apparent was electromagnetism, where already in the 1970s it was realized that the distinction between vectors and differential forms is critical for the simulation of resonance frequencies. Structure preserving integrators, which first appeared in the work of Vogalére in the 1950s[80] but whose systematic

---

[75]Marsden and Weinstein, "The Hamiltonian Structure of the Maxwell-Vlasov equations"; Marsden et al., "Hamiltonian Systems with Symmetry, Coadjoint Orbits and Plasma Physics"; Marsden, "A Group Theoretic Approach to the Equations of Plasma Physics"; Cendra et al., "The Maxwell-Vlasov Equations in Euler-Poincaré Form".

[76]Fischer et al., "General relativity as a dynamical system on the manifold $\mathcal{A}$ of Riemannian metrics which cover diffeomorphisms"; Choquet-Bruhat and Marsden, "Solution of the local mass problem in general relativity".

[77]Cf. (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, p. 2), and the quote is from (ibid., p. 116); emphasis in the original.

[78]The point of view that a physical theory is characterized by its structure is taken in much of modern physics. The governing equations, which are traditionally assumed to be the mathematical incarnation of a theory, are then derived from the structure. For example, the Euler equation for an ideal fluid is equivalent to stating that the time evolution of the fluid is a geodesic flow on the infinite dimensional Lie group $\mathrm{Diff}_\mu(\mathcal{M})$ of volume preserving diffeomorphisms of the manifold $\mathcal{M}$.

[79]Butterfield, "On Symplectic Reduction in Classical Mechanics", p. 8.

[80]Vogelaére, *Methods of integration which preserve the contact transformation property of*

study did not begin until the 1980s,[81] are another application which showed that naïvely discretizing space and time does not yield numerical algorithms which preserve desirable properties such as conserved quantities, but that respecting the geometric structure of a system is vital to obtain such techniques. The success of this work is exemplified by recent studies of the stability of the solar system—a question that concerned virtually any great mechanicist since Newton—using structure preserving numerical integrators.[82] A geometric approach to numerical simulation also provides a paradigm shift: in classical numerical analysis one is concerned with convergence in the limit when the dimensionality goes to infinity, whereas structure preserving techniques provide analogues of the mechanical system of interest which preserve its qualitative behaviour for any finite number of dimensions—the setting unavoidable for numerical computations.[83] Unsurprisingly, in many situations it is not possible to preserve the complete continuous structure in a finite setting. Nonetheless, many important properties, such as conservation of energy or time reversibility, which were and are often believed to be inherently lost in a discrete system, can be conserved in the finite setting.

2.3.1   A Primer on Geometric Mechanics[84]

Geometric mechanics is usually studied in applied mathematics and most introductory texts are hence aimed at a mathematically minded audience. In contrast, the following discussion tries to provide the intuition of geometric mechanics and to show the relevance of the subject for an understanding of "mechanics".

How does geometry get into physics?   Geometric mechanics employs modern geometry to describe mechanical systems. But how does geometry arise in mechanics? For some common mechanical systems the space of all physically possible configurations is shown in Table 2.1. For a classical particle this is just Euclidean space since its state is completely described by its position, and for a pendulum every configuration is given by the angle θ with respect to a reference axis so that the circle S^1 provides the space of all possible configurations. Already for the double pendulum, however, the situation becomes

---

the Hamiltonian equations.

[81](Ruth, "A Canonical Integration Technique"; Channell, *Symplectic integration algorithms*; Feng, "Difference schemes for Hamiltonian formalism and symplectic geometry"; Channell and Scovel, "Symplectic integration of Hamiltonian systems"; Yoshida, "Construction of higher order symplectic integrators"; Forest and Ruth, "Fourth-order symplectic integration"); on overview over the current state of the field is provided in (Hairer, Lubich, and Wanner, *Geometric Numerical Integration*) and (Marsden and West, "Discrete Mechanics and Variational Integrators").

[82]See for example (Laskar, "Large-scale chaos in the solar system"; Tremaine, "Is the solar system stable?"; Laskar and Gastineau, "Existence of collisional trajectories of Mercury, Mars and Venus with the Earth.").

[83]Grinspun, "Discrete Elastic Rods and Viscous Threads".

[84]Readers who are missing preciseness in the ensuing discussion, or are unclear about a concept that is discussed, are invited to consult Chapter 2.3.4 while reading this section.
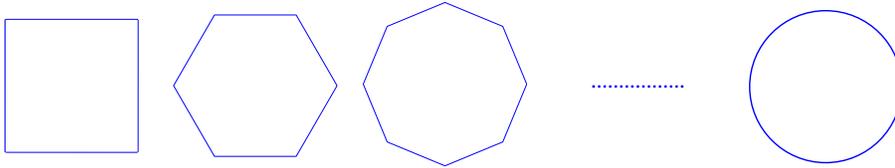
Figure 2.9: Continuous rotational symmetry as the limit of discrete rotational symmetries. Rotating a regular n-gon by $2\pi/n$ yields the same n-gon—it is preserved under the discrete set of rotations. In the continuous limit the rotation by an arbitrary angle preserves the circle S^1.

more interesting. Every configuration of the two arms is described by two angles, say θ and ϕ, and since the arms are independent of each other the space of all possible states is $S^1 \times S^1$. But the tensor product $S^1 \times S^1$ forms the torus $\mathbb{T}^2$. Instead of the arms which represent the system in physical space, we hence have an alternative representation of the system where configurations are given by points (θ, ϕ) on the "doughnut" with every point *corresponding* to a displacement of the arms. What is also apparent from the double pendulum is how constraints can be enforced intrinsically by choosing an appropriate geometric representation. The system could equivalently be described with the endpoints of the arms as particles in $\mathbb{R}^3$. But how many variables would then be needed? We would require two 3-dimensional Cartesian coordinates to describe the positions and we would need two constraints, one for the plane the pendulum lies in and one for the unit spheres on which the particles move. With the torus $\mathbb{T}^2 = S^1 \times S^1$, the natural geometric structure of the double pendulum, the two angles θ and ϕ suffice to describe all configurations. We begin to see how geometry is an intrinsic part of mechanics and why the geometry should be respected: the space of all admissible configurations of a mechanical system has a natural geometric structure and constraints are intrinsically satisfied by the choice of the geometry. In more formal parlance, the configuration space Q of a mechanical system is a manifold, the generalization of a 2-dimensional surface in space, and its topological and geometrical structure represent all physical states.

The description of the configurations of a system as points on a manifold is the principal premise of geometric mechanics and it enables to illustrate the system's structure even when the configuration space is complicated and abstract, cf. again Table 2.1, providing the inherent intuition of geometric mechanics.

What have a butterfly and a stone in common? We could end here, with manifolds as configuration spaces, and we would obtain a rich and vigorous theory. However, many mechanical systems have another and complementary geometric structure: symmetry.[85] In contrast to the discrete symmetries that might come to one's mind, such as the mirror symmetry of a butterfly or

---

[85]For a discussion of the history of the concept and its important in modern physics see (Brading and Castellani, "Symmetries and Invariances in Classical Physics").

	Physical System	Configuration space
Classical particle		
Single pendulum		
Double pendulum		
Euler top		
Ideal Euler fluid		
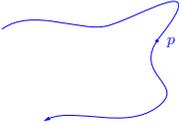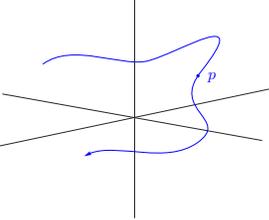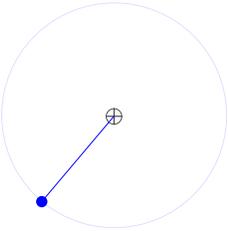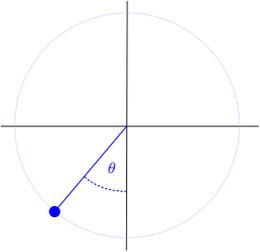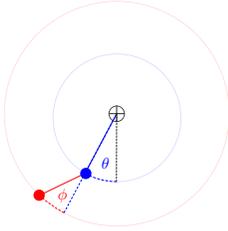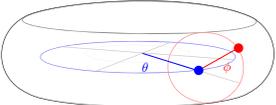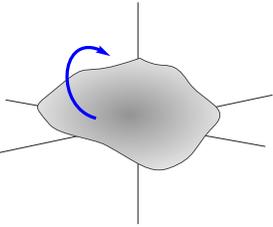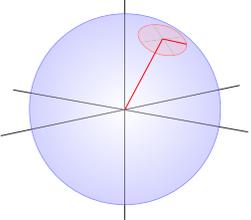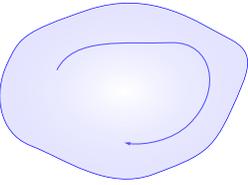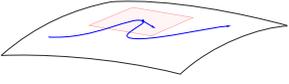
Table 2.1: Configuration space, the space of all physically valid states, for some classical mechanical systems.
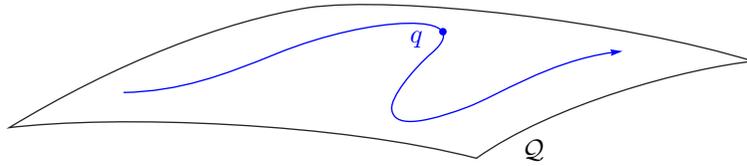
Figure 2.10: Configuration space Q of physical system with each point $q \in Q$ being a valid physical configuration of the system. Time evolution corresponds to a curve $q(t)$ on the manifold Q.

the discrete radial symmetry of flowers, mechanical system have continuous symmetries as depicted in Fig. 2.9. For example, for the classical particle considered before we can translate the coordinate system without affecting its motion—as two mirror images are equivalent so is a particle described in a translated reference frame. The single pendulum possesses a rotational symmetry where we can rotate the reference axis without changing its physical behaviour, and a similar symmetry also exists for the Euler top: a rigid body, such as a stone, that is fixed in space but free to rotate around any axis, see again Table 2.1.[86] At any time, the configuration of the Euler top is described by a rotation with respect to an initial configuration—whose geometric structure is illustrated nicely by the Poincaré map—and the continuous symmetry hence arises again from the arbitrariness of the reference configuration. Going from the 1- and 3-dimensional symmetries of the pendulum and the Euler top to an "infinite" dimensional rotational symmetry leads to the geometric structure of an ideal Euler fluid.[87] For this system, the geometry is too complicated to be visualized directly and we have to resort to an iconic representation as in Table 2.1. However, all configurations can again be described with respect to an initial reference configuration by considering the trajectories traced out by the fluid "particles", and globally this is represented by a diffeomorphism which is volume preserving since the fluid is incompressible.[88]

Formally, continuous symmetries are described by Lie groups and their action on configuration space, and their importance lies in the conserved quantities and reduced descriptions which arise from them. We will consider these aspects in more detail in the following when we discuss dynamics.

---

[86] It is named Euler top since it was Euler who first wrote down the correct equations of motion.

[87] This analogy was first pointed out by Arnold ("Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits").

[88] For a formal definition of a diffeomorphism see Definition 2.62 and also Chapter 2.3.3.3. Intuitively, one can think of a diffeomorphism as a smooth map between continuous domains that has a smooth inverse, that is every point in the first domain is smoothly mapped to a point in the second domain, and there is a map that "reverses" going from the first to the second domain.
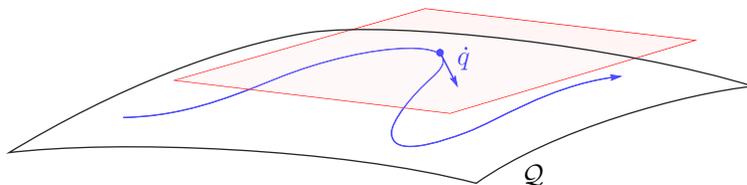
Figure 2.11: The velocity $\dot{q}(t)$ of the curve $q(t)$ lies in the tangent bundle T_qQ of the configuration manifold Q.

How does a rubber band describe the dynamics of a rigid body? We already learned that the configurations of a mechanical system are naturally described by points on its configuration manifold Q. What we are really interested in, however, are not isolated configurations but time evolution. A set of consecutive states at times $t_1 < t_2 < \ldots$ is thus given by a sequence $(q(t_1), q(t_2), \ldots)$ of locations on $\mathbb{T}^2$, see Fig. 2.12. To understand how this is described in geometric mechanics let us consider again the double pendulum. Each configuration of the system is given by two angles which represent a point $q = (\theta, \phi)$ on the torus $\mathbb{T}^2$. When the time interval Δt_i between configurations $q(t_i)$ and $q(t_{i+1})$ becomes vanishingly small, we surely expect that also the distance between the points on the torus goes to zero. But then the configurations have to form a smooth curve $q(t)$, and since the $q(t_i)$ lie on the configuration space of the double pendulum it is a curve on the torus $\mathbb{T}^2$.


Figure 2.12:
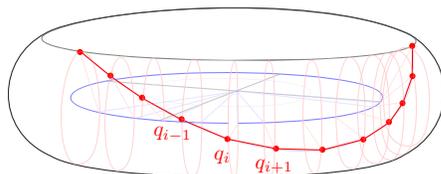
A little thought shows that our reasoning was independent of the chosen example and that for any system the time evolution can be described by a curve $q(t) : [a, b] \to Q$ on configuration space.[89] Hence, in geometric mechanics much intuition also exists for the time evolution of a system and we can illustrate it as a "marble" tracing out its path on configuration space. When configuration space is too complex to be visualized directly, such as for the Euler fluid, we will retain the geometric intuition again by considering a curve on an iconic representation as depicted in Fig. 2.10.

Given a curve $q(t) : [a, b] \to Q$ describing the time evolution of a system, we know its configuration for all times $t \in [a, b]$. Unfortunately, we rarely have this information at our disposal and all we know in most instances is the system's current state—although we still would like to determine its future configurations. Even worse, the present configuration is usually not even sufficient to determine

---

[89]For general physical systems it is not necessarily satisfied that the curve on configuration space is smooth. For example, for systems with impact the curve is in general only continuous. However, in these introductory notes we will restrict ourselves to smooth curves

Figure 2.14: A physical path between two states $\bar{q}$ and q is a stationary point of the action (blue) in the space of all possible paths on configuration space Q.

the time evolution. Fortunately, for many systems knowing its position $q(t)$ and its velocity $\dot{q}(t) = v(t) = dq(t)/dt$ provides the needed information. The geometry of the velocity vector $\dot{q}(t)$ is easily understood when we consider again the double pendulum and assume that the second pendulum is at rest, in which case the time evolution is described by a curve $q(t)$ on the equator of the torus $\mathbb{T}^2$. The velocity vector $\dot{q}(\bar{t})$ for some time $\bar{t}$ is by construction tangent to the curve. But since $q(t)$ lies on the torus $\mathbb{T}^2$, the tangent $\dot{q}(\bar{t})$ lies also in the tangent space of the manifold at the point $q(\bar{t})$ along the curve, see Fig. 2.13.[90] It is again easy to see that there was nothing special about our example, and that for any system the velocity $\dot{q}(t)$ is tangent to the curve $q(t)$ describing the time evolution of the system and it lies in the tangent space $T_{q(t)}Q$ of configuration space at $q(t)$, see Fig. 2.11 for the general picture one should have in mind.

So far we discussed how curves on configuration space enable to describe the states of a system over time. However, usually we are only given the current configuration and what we are interested in is the state in the future. A recipe for obtaining a description of the time evolution is provided by Hamilton's principle of least action. In-

Figure 2.13:

tuitively, it states that physical paths on configuration space are paths of least resistance with respect to an action functional $S(q(t))$—or the paths a rubber band would settle in on configuration space around the "hills" defined by the Lagrangian $L(q(t), \dot{q}(t))$, see Fig. 2.14 for the geometric intuition.[91] Formally, the action principle is given by

$$0 = \frac{\delta S}{\delta q(t)} = \frac{d}{d\epsilon} S(q(t) + \epsilon\, r(t)),$$
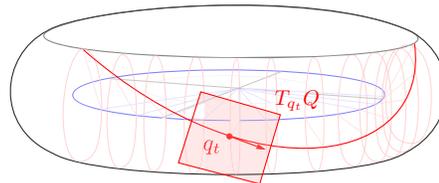
[90]As usual, the tangent space can be interpreted as the best linear approximation to the manifold at a point.

[91]The idea of least resistance is even more apparent in the Gauss-Hertz principle of least curvature (*Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*) but it is technically more involved and less general than Hamilton's principle.
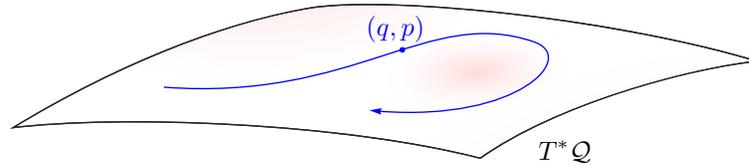
Figure 2.15: In Hamiltonian mechanics the time evolution of a mechanical system is considered on phase space $P = T^*Q$ and governed by the Hamiltonian (light red) which for each phase space point (q, p) defines the total energy of the system.

where the action S corresponds to the energy along the rubber band as defined by the Lagrangian

$$0 = \frac{\delta S}{\delta q(t)} = \frac{d}{d\epsilon} \int L\left(q(t) + \epsilon\, r(t), \frac{d}{dt}(q(t) + \epsilon\, r(t))\right) dt,$$

and it states that the paths on configuration space taken by a physical system correspond to stationary points $\delta S/\delta q(t) = 0$ of the action functional $S(q(t))$ where the functional derivative $\delta S/\delta q(t)$ vanishes. Physical trajectories are hence the local extrema of the action $S(q(t))$, analogous to the local extrema of a function over the real line which are the stationary points of the ordinary derivative. The Lagrangian $L(q(t), \dot{q}(t))$ can be understood as the characteristic function of a system—depending both on its configuration $q(t)$ and its velocity $\dot{q}(t)$—and it is usually defined as the kinetic minus the potential energy of a system. For example, for a classical particle of mass m in a potential $V(q)$ the Lagrangian is

$$L(q, \dot{q}) = \frac{m}{2}\|\dot{q}\|^2 - V(q).$$

Using the calculus of functional derivatives, one can derive from Hamilton's principle differential equations describing the motion of the system. In the general case, the equations are known as Euler-Lagrange equations, and for the above Lagrangian these are equivalent to Newton's equations of motion.

Next to the action principle, an alternative way to describe the dynamics of a mechanical system is Hamiltonian mechanics where the time evolution is governed by the system's energy. Instead of using the velocity $\dot{q} \in TQ$ which determines the change in a system's configuration, it is then useful to employ a description of the change in the system's (kinetic) energy T. This is conveniently expressed using a co-vector p in the dual space T^*Q "measuring" the change δT with the pairing $p(\dot{q}) = p \cdot \dot{q}$. For the classical particle we considered before, the co-vector is given by $p = m\,\dot{q}$ where the mass m expresses the dependence on the kinetic energy $T = 1/2\, m\, \|\dot{q}\|^2$, and p is hence what Leibniz and Newton called momentum.[92] With the momentum, an alternative description of a

---

[92]Incidentally, we also have $p(\dot{q}) = p \cdot \dot{q} = m\,\dot{q} \cdot \dot{q} = m\, \|\dot{q}\|^2 = 2\,T$. But didn't we say the
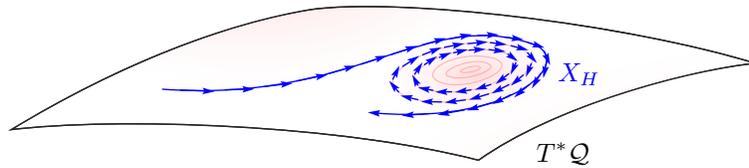
Figure 2.16: Flow on phase space along the Hamiltonian vector field X_H (blue arrows) preserving the level sets of the Hamiltonian (light red).

system is possible by "lifting" it to the cotangent bundle T^*Q and on this 'phase space' a configuration is then given by the system's position q as well as its momentum p, see Fig. 2.15. Evidently, the lift makes a system's description more complex in that the number of variables describing a configuration is doubled. However, as might already appear reasonable from intuition, it enables a "simpler" description of the dynamics. Using a similar reasoning as before, it is not hard to see that on T^*Q time evolution is again represented by a smooth curve $z(t) = (q(t), p(t)) : [a, b] \to T^*Q$. However, in contrast to Lagrangian mechanics on configuration space where Hamilton's action principle determines trajectories, the evolution of a point $z(t) = (q(t), p(t))$ along the curve on phase space is governed by a simple law: the total energy of the system given by the Hamiltonian $H(q, p) : T^*Q \to \mathbb{R}$ is conserved.[93] This leaves us again with the question how we can determine the curve describing the time evolution from an initial configuration? To find the answer, let us return to the classical particle in a potential. The Hamiltonian for the system is given by

$$H(q, p) = \frac{\|p\|^2}{2\,m} + V(q)$$

and its phase space is $\mathbb{R}^3 \times \mathbb{R}^3$ since the tangent space of Euclidean space $\mathbb{R}^3$ is the space itself and since we can think of the momentum as a vector in $\mathbb{R}^3$.[94] But in Euclidean space it is not hard to find the direction where a function $f : \mathbb{R}^3 \to \mathbb{R}$ does not change: the gradient ∇f determines the direction of maximal change and the change in an arbitrary direction $\vec{u}$ is given by the dot product $\nabla f \cdot \vec{u}$, and hence any vector orthogonal to the gradient direction ∇f defines a direction where the value of f is conserved. With the gradients with respect to position and momentum being ∇_q and ∇_p, respectively, the tangent vector X_H on phase space defining the direction where the Hamiltonian does

---

momentum determines the *change* in the kinetic energy? How can p at the same time provide the change *and* the total energy?

[93]For some systems the Hamiltonian does not directly represent the total energy of a system but these are far beyond the scope of our considerations.

[94]In most applications it is useful to *not* identify $T\mathbb{R}^n$ with $\mathbb{R}^n$ but to carefully distinguish the two spaces, and it is similarly usually not advisable to identify TQ and its dual T^*Q even if this possible using a metric. However, for the simplicity of our argument we will employ the usual obfuscations.

not change is given by

$$X_H = \mathbb{J}\,dH \cong \begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \nabla_q H \\ \nabla_p H \end{pmatrix} = \begin{pmatrix} \nabla_p H \\ -\nabla_q H \end{pmatrix}$$

where I denotes the 3×3 identity matrix and the above equations are known as Hamilton's equations. The symplectic matrix $\mathbb{J}$ indeed ensures that the Hamiltonian $H(q,p)$ is conserved along the flow of the Hamiltonian vector field X_H since

$$dH\,(X_H) = \nabla H \cdot X_H = \nabla_q H \cdot \nabla_p H - \nabla_p H \cdot \nabla_q H = 0$$

and although the flow is defined on 6-dimensional phase space much intuition about its behaviour can be gained by considering the underlying geometry as depicted in Fig. 2.16. Needless to say, the above conception of the dynamics of Hamiltonian mechanics applies to any phase space when a generalized gradient and an intrinsic definition of the symplectic matrix are employed.

Lagrangian and Hamiltonian mechanics provide alternative descriptions of a mechanical system, and when they are equivalent, as is usually the case, one can change the point of view using the Legendre transform. However, each perspective also provides its own merits and demerits, and often one of the descriptions appears more natural. For example, as we saw before, Lagrangian mechanics is defined on configuration space Q while Hamiltonian mechanics employs "lifted" dynamics on phase space $P = T^*Q$. This leads to second order differential equations for the time evolution in the Lagrangian picture and to first order equation in the Hamiltonian; more concretely, Newton's equations, which are equivalent to the Euler-Lagrange equation for a Lagrangian of the form considered before, depend on the acceleration, the second derivative of position with respect to time, while in Hamilton's equations as introduced above the Hamiltonian vector field X_H is the first time derivative of position and momentum—it is this reduction from second to first order differential equations which provides the "simplification" in Hamiltonian dynamics which we advertised before.

So far we did not consider symmetries when we described the time evolution of mechanical systems. Nonetheless, they are a vital aspect of geometric mechanics since they allow to restrict a system's dynamics to the level sets of the conserved quantities, see Fig. 2.17. This is possible by Noether's theorem which assures us that continuous symmetries lead to conserved quantities invariant under the dynamics. We already encountered one such quantity: the Hamiltonian or energy of a system. Conservation of energy is associated with invariance under time translation, and it shows that we tacitly assumed the time invariance of the Hamiltonian in the foregoing. Other conserved quantities that are often encountered are linear and angular momentum which are associated with translational and rotational invariance, respectively, symmetries we saw before for the classical particle and the pendulum. Restricting the dynamics of a system to the level sets of the conserved quantities is known as reduction and for many systems critical to obtain an effective description of its time evolution.
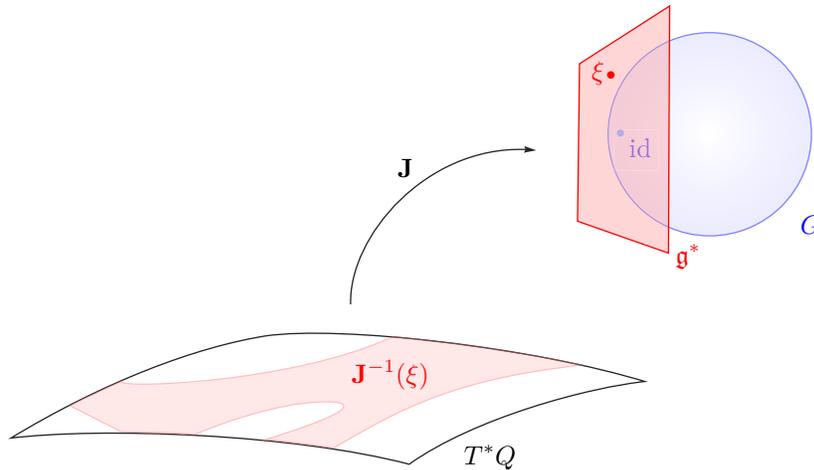
Figure 2.17: Symmetries allow to restrict phase space T^*Q to the level sets $\mu^{-1}(\xi)$ of the conserved quantities which are elements ξ in the dual Lie algebra $\mathfrak{g}^*$. The connection between phase space and the dual Lie algebra is established by the momentum map μ^* which provides the modern Hamiltonian formulation of Noether's theorem.

Old wine in new skins? We have seen that manifolds arise naturally in the description of classical mechanical systems by providing the space of all possible configurations, and that many constraints are enforced intrinsically by the topology and shape of the geometry. Time evolution in geometric mechanics is represented by curves on the configuration manifold, and we outlined how it can be determined using Lagrangian and Hamiltonian mechanics.

An important trait of geometric mechanics is its intuitive nature. The structure and time evolution of a mechanical system can be illustrated by visualizing configuration and phase space, which as we have done on the preceding pages is possible even if the spaces are complex and high dimensional, and practically working with geometric mechanics often means to exploit this inherent geometric intuition.[95] It goes without saying that formal mathematics has its place in geometric mechanics by complementing intuition and making it rigorous, and that the mastery of modern differential geometry and tensor analysis is essential when geometric mechanics is employed to study mechanical systems. Next to the intuitive appeal, a second characteristic of geometric mechanics is its emphasis on mathematical and physical structure. While Newtonian mechanics is highly descriptive, making it easy to learn and to carry out computations, it does not reveal structure. In geometric mechanics, in contrast, computations are structural arguments which provide insight into the fabric they represent, and this structural insight explains much of the vigour of geometric mechanics.[96]

---

[95]The intuitionist approach in geometric mechanics is to be understood in the naïve sense of the word, and not in the sense of Brouwer's programme.

[96]Lanczos described this as: "Since the days of antiquity it has been the privilege of the

In our discussion we only considered classical mechanical systems. However, the theory applies to a diverse array of fields and disciplines ranging from quantum mechanics at the smallest scales to relativistic astrophysics at the largest, and applications can be found in areas such as image processing, space mission design, marine animal propulsion, mathematical finance, rising eggs, oceanography, plasma physics, falling cat phenomena, and many more. In its contemporary formulation using the rich toolbox of modern geometry, geometric mechanics provides thereby a surprisingly unified perspective on all these systems.[97] Although we considered them only cursory, symmetries are an integral part of geometric mechanics and much of the theory is devoted to reduction theory: obtaining simpler descriptions of systems by exploiting symmetries and conserved quantities.

In the next section we will introduce the formal mathematics that is needed for geometric mechanics, and afterwards the intuitive perspective of the theory presented in this section will be made rigorous. Nonetheless, even when a more formal language is employed than in our primer, it should be kept in mind that geometric mechanics is an intuitive endeavour and that pictures are the key to understanding.

2.3.2 Calculus on Manifolds

In the following we will formally introduce the mathematical formalism required for geometric mechanics. Our treatment will be more general than that found in many texts since we require infinite dimensional manifolds, which at times leads to considerable complications in the theory. However, following the literature we will avoid functional analytic questions most of the time and treat certain aspects formally, for example without studying convergence. Elementary notions of point set topology and differential calculus will be assumed,[98] and knowledge of vector calculus in three dimensions will be useful. Here and in the following we will assume that all objects are sufficiently smooth for the operations of interest to us.

---

mathematician to engrave his conclusions, expressed in a rarefied and esoteric language, upon the rocks of eternity. While this method is excellent for the codification of mathematical results, it is not so acceptable to the many addicts of mathematics, for whom the science of mathematics is not a logical game, but the language in which the physical universe speaks to us, and whose mastery is inevitable for the comprehension of natural phenomena.", (Lanczos, *Linear Differential Operators*, p. vii).

[97]Marsden and Ratiu expressed the unified perspective provided by geometric mechanics as follows: "Even more striking are true statements like this: 'Don't tell me that quantum mechanics is right and classical mechanics is wrong—after all, quantum mechanics is a special case of classical mechanics.", (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, p. 116).

[98]See for example (Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*, Chapter 1 and 2) for the required material.

2.3.2.1   Preliminaries

Derivative   Before we can define smooth objects such as manifolds a sufficiently general notion of differentiability is needed.

Definition 2.56. *Let* $\mathbf{E}, \mathbf{F}$ *be normed vector spaces and* $f : U \subset \mathbf{E} \to \mathbf{F}$ *for some open set* U. *The map* f *is said to be* ***differentiable*** *at* $\bar{u} \in U$ *if a bounded linear map* $\mathrm{D}f_{\bar{u}} : \mathbf{E} \to \mathbf{F}$ *exists such that for every* $\epsilon > 0$ *there is a* $\delta > 0$ *such that for* $u \in U$ *satisfying* $0 < \|u - \bar{u}\| < \delta$ *implies*

$$\frac{\|f(u) - f(\bar{u}) - (\mathrm{D}f_{\bar{u}}) \cdot (\bar{u} - u)\|}{\|u - \bar{u}\|} < \epsilon,$$

where $(\mathrm{D}f_{\bar{u}}) \cdot e$ *denotes the evaluation of the map* $\mathrm{D}f_{\bar{u}}$ *at* $e \in \mathbf{E}$. *If* f *is differentiable at every* $u \in U$ *the map* $\mathrm{D}f : U \to L(\mathbf{E}, \mathbf{F})$ *given by* $u \to \mathrm{D}f$ *is the* ***derivative*** *of* f.

In the above definition, $L(\mathbf{E}, \mathbf{F})$ is the space of linear operators from $\mathbf{E}$ to $\mathbf{F}$. Intuitively, Def. 2.56 generalizes the idea that the derivative $f'(\bar{x})$ of a scalar function f provides the tangent $y = f'(\bar{x})(x - \bar{x}) = m_{\bar{x}}(x - \bar{x})$ which is the best linear approximation of f as a functional acting on $(x - \bar{x})$. It can be shown that, if it exists, $\mathrm{D}f$ is unique and satisfies the usual properties of the derivative such as the Leibniz rule and the chain rule.

Remark 2.40. Let $f : U \subset R^n \to \mathbb{R}^m$ be differential and given in coordinates by $f(x^1, \ldots, x^n) = (f^1(x^1, \ldots, x^n), \ldots, f^m(x^1, \ldots, x^n))$. The matrix representation J_f of the linear map $(\mathrm{D}f)(x)$ is then the **Jacobian matrix**

$$J_f(x) = \begin{bmatrix} f^1/x^1 & \cdots & f^1/x^n \\ \vdots & \ddots & \vdots \\ f^m/x^1 & \cdots & f^m/x^n \end{bmatrix}.$$

A special derivative of much importance is the following.

Definition 2.57. *Let* $f : U \subset \mathbf{E} \to \mathbf{F}$ *be a map between normed vector spaces with* $U \subset \mathbf{E}$ *open. The* ***directional derivative of*** f ***at*** u, *if it exists, is*

$$\left. \frac{d}{dt} f(u + t\, e) \right|_{t=0}.$$

The directional derivative as defined above is sometimes known as Fréchet derivative. The following theorem connects the directional derivative with the derivative.

Proposition 2.40. *Let* $f : U \subset \mathbf{E} \to \mathbf{F}$ *be a map between normed vector spaces with* $U \subset \mathbf{E}$ *open. If* f *is differentiable at* u *then the directional derivatives exist and is given by*

$$\left. \frac{d}{dt} f(u + t\, e) \right|_{t=0} = \mathrm{D}f(u) \cdot e.$$
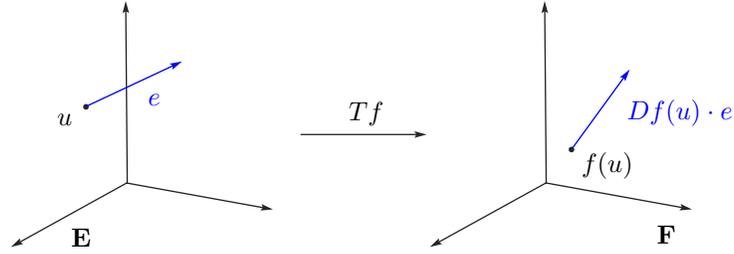
Figure 2.18: Tangent map $Tf(u, e) = (f(u), \mathrm{D}f(u) \cdot e)$ between linear spaces $\mathbf{E}$ and $\mathbf{F}$.

If $f : \mathbb{R}^n \to \mathbb{R}$ is a map from $\mathbb{R}^n$ into $\mathbb{R}$ then the directional derivative takes the usual form known from vector calculus, cf. Def. 2.60 below.

Definition 2.58. *Let $\mathbf{E}, \mathbf{F}$ be normed vector spaces and $f, g : U \subset \mathbf{E} \to \mathbf{F}$. The maps f and g are **tangent** at $\bar{u} \in U$ if $f(\bar{u}) = f(\bar{u})$ and*

$$\lim_{u \to \bar{u}} \frac{\|f(u) - g(u)\|}{\|u - \bar{u}\|} = 0.$$

The above definition makes precise the idea that two mappings f, g agree to first order at some point $\bar{u} \in U$.

Definition 2.59. *Let $f : U \subset \mathbf{E} \to \mathbf{F}$ be differentiable for an open set $U \subset \mathbf{E}$, and let $\mathbf{E}, \mathbf{F}$ be normed vector spaces. The **tangent** $Tf : U \times \mathbf{E} \to \mathbf{F} \times \mathbf{F}$ of f is the map*

$$(Tf)(u, e) = (f(u), (\mathrm{D}f) \cdot e).$$

See Fig. 2.18. The significance of the spaces $U \times \mathbf{E}$ and $\mathbf{F} \times \mathbf{F}$ will become clear when we introduce tangent spaces. The tangent satisfies the chain rule in that $T(f \circ g) = Tf \circ Tg$ for f and g being suitable mappings.

For continuum mechanics we will also need a functional derivative that allows to differentiate functionals with respect to functions. The definition extends the concept of the gradient which will be introduced first.

Definition 2.60. *Let $\mathcal{H}$ be a Hilbert space and $f : U \subset \mathcal{H} \to \mathbb{R}$ for some open set U. The **differential** of f is the element $\mathrm{d}f \in \mathcal{H}^*$ in the dual space $\mathcal{H}^*$ such that $\mathrm{d}f(e) = \mathrm{D}f \cdot e$. The **gradient** ∇f is the map $\nabla f : U \to \mathcal{H}$ defined by $\langle \nabla f, e \rangle = \mathrm{d}f \cdot e$.*

The differential is an element in the dual space $\mathcal{H}^*$ since for the map $f : U \subset \mathcal{H} \to \mathbb{R}$ the range is $F = \mathbb{R}$, cf. Def. 2.56, but by definition the dual space is the space of functionals $F : \mathcal{H} \to \mathbb{R}$, cf. Def. 2.18. The definition of the gradient relies on the Riesz representation theorem, Theorem 2.5, that enables to identify $\mathcal{H}$ and $\mathcal{H}^*$, and the usual gradient is recovered with $\mathcal{H} = \mathbb{R}^3$. A functional derivative, usually denoted by $\delta F / \delta u$, is obtained when $\mathcal{H}$ is a function space. Using the concept of a weakly non-degenerate pairing introduced in

Def. 2.29, a slightly more general notion of functional derivative can be obtained which is available in general Banach spaces.

Definition 2.61. *Let* $\mathbf{E}$ *and* $\mathbf{F}$ *be Banach spaces and* $\langle\,,\rangle$ *an* $\mathbf{E}$*-weakly non-degenerate pairing between the spaces. The* **functional derivative** $\delta F/\delta\varphi$ *of the differentiable map* $F : \mathbf{F} \to \mathbf{E}$ *at* $\varphi \in \mathbf{F}$ *is the element in* $\mathbf{E}$ *such that for all* $\psi \in \mathbf{F}$ *it holds*

$$\mathrm{D}F(\varphi) \cdot \psi = \left\langle \frac{\delta F}{\delta\varphi}, \psi \right\rangle.$$

The weak non-degeneracy of the pairing ensures that the functional derivative is unique, if it exists. As for the gradient, we employ the pairing $\langle\,,\rangle$ to identify $\mathrm{D}F$ with an element in $\mathbf{E}$ although uniqueness can no longer be guaranteed. Using Proposition 2.40, the functional derivative is usually determined by

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} F(\varphi + \epsilon\psi) = \left\langle \frac{\delta F}{\delta\varphi}, \psi \right\rangle \tag{2.69}$$

and the left hand represents a first order ϵ-perturbation of the functional F at φ by the function ψ.

Example 2.32. The calculus of variations is a special case of the functional derivative where functionals defined over the (infinite dimensional) space of all paths $q(t)$ on a space are considered, cf. Fig. 2.14. The action functional, whose derivative is sought, is defined by

$$\mathcal{S}(q(t)) = \int_\Omega L\left(q(t), \dot{q}(t)\right)\, dt, \tag{2.70}$$

where $\Omega = [0, T] \subset \mathbb{R}$, $\dot{q}(t)$ denotes the time derivative of $q(t)$, and $L(q(t), \dot{q}(t))$ is a smooth function on the space over which the paths are defined. With the ansatz in Eq. 2.69 and the pairing being given by integration, we have

$$\frac{\delta S}{\delta q(t)} = \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} S(q(t) + \epsilon\, r(t)) \tag{2.71a}$$

$$= \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \int_\Omega L\left(q(t) + \epsilon\, r(t), \frac{d}{dt}\left(q(t) + \epsilon\, r(t)\right)\right) dt. \tag{2.71b}$$

By the linearity of the derivative, and writing $q_\epsilon(t) = q(t) + \epsilon\, r(t)$, the total derivative with respect to ϵ is

$$\frac{\delta S}{\delta q(t)} = \int_\Omega \frac{\partial L}{\partial q_\epsilon(t)} \frac{\partial q_\epsilon(t)}{\partial\epsilon} + \frac{\partial L}{\partial \dot{q}_\epsilon(t)} \frac{\partial \dot{q}_\epsilon(t)}{\partial\epsilon}\, dt \bigg|_{\epsilon=0}. \tag{2.71c}$$

Exploiting that partial derivatives commute yields

$$\frac{\delta S}{\delta q(t)} = \int_\Omega \frac{\partial L}{\partial q_\epsilon(t)} \frac{\partial q_\epsilon(t)}{\partial\epsilon} + \frac{\partial L}{\partial \dot{q}_\epsilon(t)} \frac{d}{dt} \frac{\partial q_\epsilon(t)}{\partial\epsilon}\, dt \bigg|_{\epsilon=0} \tag{2.71d}$$

and using integration by parts for the second term we obtain

$$\frac{\delta S}{\delta q(t)} = \int_\Omega \frac{\partial L}{\partial q_\epsilon(t)} \frac{\partial q_\epsilon(t)}{\partial \epsilon} \, dt \bigg|_{\epsilon=0} + \int_\Omega \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\epsilon(t)} \frac{\partial q_\epsilon(t)}{\partial \epsilon} \, dt \bigg|_{\epsilon=0}$$
$$+ \left[\frac{\partial L}{\partial \dot{q}_\epsilon(t)} \frac{\partial q_\epsilon(t)}{\partial \epsilon} \right]_{t=0}^{T} \bigg|_{\epsilon=0}. \tag{2.71e}$$

The boundary term vanishes if one requires that $r(t)$ vanishes at the endpoint $\{0, T\}$. Taking into account that we are interested in $\epsilon = 0$ and resolving $\partial q_\epsilon(t)/\partial \epsilon$ yields

$$\frac{\delta S}{\delta q(t)} = \int_\Omega \frac{\partial L}{\partial q(t)} \, r(t) \, dt + \int_\Omega \frac{d}{dt} \frac{\partial L}{\partial \dot{q}(t)} \, r(t) \, dt \tag{2.71f}$$

$$= \int_\Omega \left(\frac{\partial L}{\partial q(t)} + \frac{d}{dt} \frac{\partial L}{\partial \dot{q}(t)} \right) r(t) \, dt. \tag{2.71g}$$

The last equation is in the form of the right hand side of Eq. 2.69 and the functional derivative is hence

$$\frac{\delta S}{\delta q(t)} = \frac{\partial L}{\partial q(t)} + \frac{d}{dt} \frac{\partial L}{\partial \dot{q}(t)}. \tag{2.72}$$

If one is interested in stationary points of the functional, to first order as in our ansatz in Eq. 2.69, one obtains, in analogy with local extrema of ordinary functions, that

$$\frac{\partial L}{\partial q(t)} + \frac{d}{dt} \frac{\partial L}{\partial \dot{q}(t)} = 0 \tag{2.73}$$

has to be satisfied. The last equation is known as Euler-Lagrange equation and it is central to mechanics. For an appropriate Lagrangian $L(q, \dot{q})$ it is equivalent to Newton's second law but it generalizes for example to electromagnetic theory.

A mapping which requires the notion of a derivative and which will play an outstanding role in the following is a diffeomorphism. We will introduce it next.

Definition 2.62. *A **diffeomorphism** $\varphi : \mathcal{A} \to \mathcal{B}$ between Banach spaces $\mathcal{A}$ and $\mathcal{B}$ is a bijective map that is smooth and has a smooth inverse.*

The Inverse and Implicit Function Theorems Two results of central importance in nonlinear analysis are the inverse and implicit function theorems. Although we do not employ them directly, they are part of the fabric of the theory and used in many of the proofs of the results we are employing.

Theorem 2.11 (Inverse Function Theorem)**.** *Let $\mathbf{E}, \mathbf{F}$ be Banach spaces and $\varphi : U \subset \mathbf{E} \to \mathbf{F}$ be a smooth mapping from an open subset $U \subset \mathbf{E}$ to $\mathbf{F}$. For $\bar{x}$ let $\mathrm{D}\varphi(\bar{x})$ be a linear isomorphism. Then φ is a local diffeomorphism of some*

neighborhood $\bar{U}$ of $\bar{x}$ onto some neighborhood $\bar{V}$ of $\varphi(\bar{x})$ and the derivative of the inverse function φ^{-1} at $y \in \bar{V}$ is given by

$$\mathrm{D}\varphi^{-1}(y) = \left[\mathrm{D}\varphi(\varphi^{-1}(y))\right]^{-1}.$$

In the context of infinite-dimensional Banach spaces, the theorem is also known as inverse mapping theorem. Estimates of the size of the neighborhoods on which the theorem holds are also possible, for example using second derivatives. The inverse function theorem enables, at least locally, to transfer results from **E** to **F** and vice versa.

The implicit function theorem is employed to express an implicit relation as an explicit function, which makes the usual tools from functional analysis available.

Theorem 2.12 (Implicit Function Theorem). *Let $U \subset \mathbf{E}$ and $V \subset \mathbf{F}$ be open neighborhoods of Banach spaces and $\varphi : U \times V \to \mathbf{G}$ be smooth, where $\mathbf{G}$ is also a Banach space. Moreover, for $\bar{x} \in U$, $\bar{y} \in V$ let the derivative $D_y\varphi(\bar{x}, \bar{y})$ of $\varphi : \mathbf{F} \to \mathbf{G}$ with respect to the second argument be an isomorphism. Then there are neighborhoods $\bar{U}$ of $\bar{u}$ and $\bar{W}$ of $\varphi(\bar{x}, \bar{y})$ and a unique map $g : \bar{U} \times \bar{W} \to V$ such that for all $(x, z) \in \bar{U} \times \bar{W}$ one has*

$$\varphi(x, g(x, z)) = z$$

and the graph $(x, \varphi(x, z))$ satisfies $\varphi(x, y)$.

Example 2.33. Let $\varphi(x, y) = \sqrt{x^2 + y^2} = 1$ be the implicit function defining the circle S^1. Clearly, there is no globally defined function $g(x) : \mathbb{R} \to \mathbb{R}$ such that the graph $(x, y = g(x))$ satisfies $\varphi(x, y) = \varphi(x, g(x)) = 1$. However, locally for every $\bar{x}$ there exists a neighborhood $x \in \bar{U} \subset [-1, 1]$ such that $g(x) = \pm\sqrt{1 - x^2}$.

2.3.2.2 Manifolds

Manifolds A manifold is the generalization of a curved surface in three dimensional space that can be defined without reference to an embedding space.[99] Each neighborhood of a manifold looks locally like a linear space, and in the finite dimensional case locally like Euclidean space. The pre-image of the local homeomorphism for each neighborhood is a chart, and different charts are required to smoothly fit together, cf. Fig. 2.19.

Definition 2.63. *Let S be a set and $\mathbf{E}$ be a Banach space. A **chart** (U, φ) on S is a bijection φ from a subset $U \subset S$ to an open subset of the **model space** $\mathbf{E}$ for (U, φ). An **atlas** $\mathcal{A}$ is a family of charts $\mathcal{A} = \{(U_1, \varphi_1), \ldots\}$ such that $S = \bigcup U_i$ and the U_i form a cover of S, and the charts are compatible in their overlap in that the **transition maps** $\varphi_{ij} = \varphi_i \circ \varphi_j^{-1} \mid_{\varphi_j(U_i \cap U_j)}$ are diffeomorphisms between the open sets $\varphi_j(U_i \cap U_j)$ and $\varphi_i(U_i \cap U_j)$. A chart is an **admissible chart** if it is part of an atlas.*

---

[99]It is a well known result by Whitney that, nonetheless, any manifold can be embedded in a high dimensional Euclidean space.
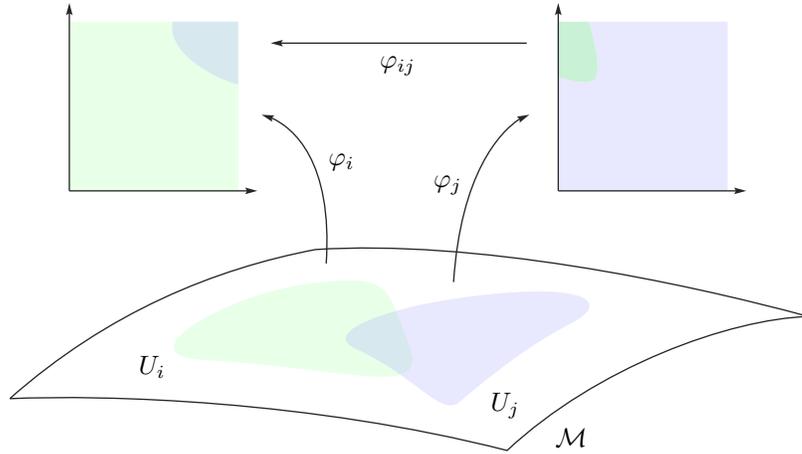
Figure 2.19: A manifold $\mathcal{M}$ with charts (U_i, φ_i) and (U_j, φ_j). On the overlap the charts are required to fit together so that the transition map $\varphi_{ij} = \varphi_i \circ \varphi_j^{-1}$ are local diffeomorphisms.

In practice, often the pre-image $\varphi_i^{-1}(U)$ is denoted as chart. By Def. 2.56, the notion of differentiability for the maps φ_{ij} is well defined since they are maps between open sets $\varphi_j(U_i \cap U_j)$ and $\varphi_i(U_i \cap U_j)$ of Banach spaces, and the compatibility requirement for the transition maps ensures that calculations are independent of the chart that is used. If the transition maps are C^k then the atlas is known as a C^k atlas. Charts are the principal tool to generalize notions from linear spaces to the nonlinear setting of manifolds: employ a local definition in the chart (U_i, φ_i) and "transfer" the concept onto the set S using the inverse chart map φ^{-1} while ensuring that the result is independent of the choice of (U_i, φ_i).

Definition 2.64. *Two atlases $\mathcal{A}_1$ and $\mathcal{A}_2$ for a set S are **equivalent** if the union $\mathcal{A}_1 \cup \mathcal{A}_2$ is again an atlas for S. If $\mathcal{A}$ is an atlas for S then the union of all atlases equivalent to $\mathcal{A}$ is the **differentiable structure** $\mathcal{D}$ generated by $\mathcal{A}$.*

The above definitions are important for a manifold to be independent of the specific choices made for an atlas.

Definition 2.65. *A **differentiable manifold** $\mathcal{M}$ is a tuple $(S, \mathcal{D})$ of a set S and a differentiable structure $\mathcal{D}$ for S.*

For notational convenience, the manifold $\mathcal{M}$ is usually identified with the underlying set S and one just speaks of $\mathcal{M}$ when one speaks of the tuple $(S, \mathcal{D})$. In the literature, often only the finite dimensional case is considered but we will need the more general notion of an infinite dimensional manifold as configuration space for continuum mechanics. The restriction to Banach spaces, instead of more general linear spaces such as Fréchet space that are often considered
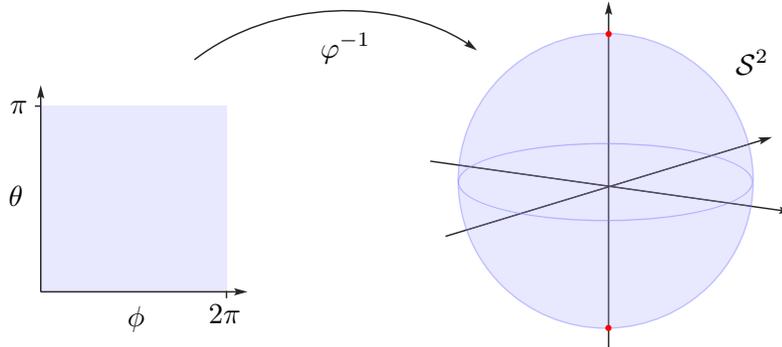
Figure 2.20: The sphere S^2 with a chart provided by spherical coordinates (θ, ϕ). The two poles (red) are not part of the chart and it hence does not provide a cover for S^2.

in mathematics, avoids the technical difficulties that arise when the inverse function theorem is no longer available. In the finite dimensional case the model spaces are Euclidean space $\mathbb{R}^n$ and the coordinate maps are homeomorphisms. Note that the definition of a manifold does not make any assumption on the number of connected components and even their dimensionality can differ.

Remark 2.41. Let $\mathcal{M}$ be a n-dimensional manifold and (U, φ) be a chart of $\mathcal{M}$ with $\varphi : U \to V \subset \mathbb{R}^n$. Then for every $u \in U$ there are coordinates $\varphi(u) = (x^1(u), \dots, x^n(u))$ in the model space V. For fixed index i the $x^i(u)$ provide maps $x^i : U \to V : u \to x^i(u)$ and these are the **local coordinates** of $\mathcal{M}$ defined by the chart (U, φ).

Example 2.34. Euclidean space $\mathbb{R}^3$ is a manifold that can be covered by a single chart (U, φ) with $U = \mathbb{R}^3$ and $\varphi = \mathrm{id}$. An alternative chart is provided by spherical coordinates although these do not provide a cover for $\mathbb{R}^3$ since the origin is a singular point of the parametrization.

The above example is easily generalized to arbitrary linear spaces: a linear space can always be covered by a single chart with U being the model space and φ the identity map. Note that $\mathbb{R}^3$ is a priori not embedded into an "ambient" space.

Example 2.35. The 2-sphere S^2 is arguably the simplest manifold that exhibits the complications of the general theory. Nonetheless, for S^2 many results are well known and have been developed using classical vector calculus, making a comparison with the modern theory interesting.

Considered as an embedded submanifold in $\mathbb{R}^3$ a chart (U, φ) for S^2 is given by spherical coordinates (θ, ϕ) with

$$\varphi(U) = \left\{ (\theta, \phi) \mid (\theta, \phi) \in (0, \pi) \times [0, 2\pi] \subset \mathbb{R}^2 \right\}, \tag{2.74a}$$

and an inverse chart map

$$\varphi^{-1} = \left(\begin{array}{c} x(\theta,\phi) \\ y(\theta,\phi) \\ z(\theta,\phi) \end{array} \right) = \left(\begin{array}{c} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{array} \right) : \mathbb{R}^2 \to \mathbb{R}^3. \qquad (2.74b)$$

The model space for S^2 is thus $\mathbb{R}^2$ and the inverse chart map is given by the well known expression for spherical coordinates, see also Fig. 2.20. However, the chart (U,φ) does not provide an atlas for S^2 since φ is singular at the poles and a neighborhood of the poles is hence not homeomorphic to a neighborhood of $\mathbb{R}^2$ under φ, although this singularity is often disregarded in applications since it is a set of measure zero that does not affect integral values over S^2. An atlas for the sphere can be constructed using two spherical coordinate systems that are restricted to not include the poles and oriented with respect to each other to cover all of S^2. An alternative to spherical coordinates is provided by stereographic projections. However, again two charts are needed to cover all of S^2 without a singularity and to provide an atlas, and it is well known that in fact always at least two charts are needed to cover S^2.

Definition 2.66. *The **neighborhood** of a point $m \in \mathcal{M}$ on a manifold $\mathcal{M}$ is the inverse image $\varphi^{-1}(N(\varphi(m)))$ of the vector space neighborhood $N(\varphi(m))$ of a point $\varphi(m)$ in an admissible chart.*

In the following we will assume that two different points $m, \bar{m} \in \mathcal{M}$ have non-intersecting neighborhoods, and the manifold is hence a Hausdorff space. The requirement is important for example for the existence of diffeomorphisms of manifolds.[100]

Definition 2.67. *Let $\mathcal{M}$ be a manifold with model space $\mathbf{E} = \mathbf{F}_1 + \mathbf{F}_2$. A **submanifold** B of $\mathcal{M}$ is a subset $B \subset \mathcal{M}$ such that for every $b \in B$ there is an admissible chart (U,φ) of $\mathcal{M}$ with $b \in U$ such that the submanifold property is satisfied:*

$$\varphi : U \to \mathbf{F}_1 \times \mathbf{F}_2 \quad \text{with} \quad \varphi(U \cap B) : U \to \mathbf{F}_1 \times \{0\}.$$

The definition makes precise the intuitive idea that a submanifold B is "generated" by a subset of the model space of $\mathcal{M}$ for each chart overlapping B. Linear transformations are ubiquitous in the theory of vector spaces. The analogue for manifolds will be introduced next.

Definition 2.68. *Let $\mathcal{M}, \mathcal{N}$ be manifolds with model spaces $\mathbf{E}$ and $\mathbf{F}$, and let $f : \mathcal{M} \to \mathcal{N}$. The **map** f **is smooth** if for every $x \in \mathcal{M}$ and $f(x) \in \mathcal{N}$, and charts (U,φ) with $x \in U$ and (V,ψ) with $f(x) \in V$, the local representative $f_{\psi\varphi} = \psi \circ f \circ \varphi^{-1} : \varphi(U) \subset \mathbf{E} \to$ is smooth.*

The local representative $f_{\psi\varphi}$ of f is a map between normed linear space so that differentiability is well defined by Def. 2.56, cf. Fig. 2.21. A smooth map

---

[100]See for example (Arnold, *Ordinary Differential Equations*, §35.2).
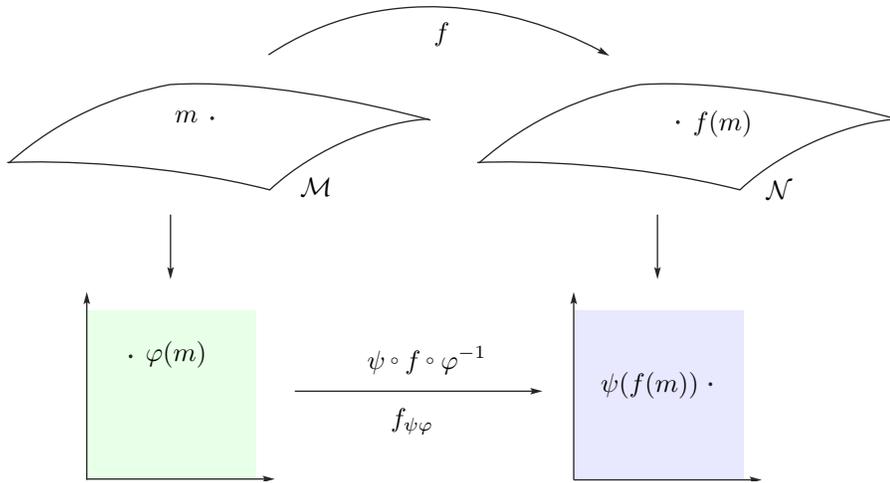
Figure 2.21: A map $f : \mathcal{M} \to \mathcal{N}$ between manifolds is realized through its local representative $f_{\psi\varphi} = \psi \circ f \circ \varphi^{-1}$ that maps between charts.

$f : \mathcal{M} \to \mathcal{N}$ whose inverse is also smooth provides a diffeomorphism between the manifolds, cf. Def. 2.62, and such a map is an isomorphism in the category of smooth manifolds.

Remark 2.42. Let $\mathcal{M}, \mathcal{N}$ be manifolds with a diffeomorphism $f : \mathcal{M} \to \mathcal{N}$, and (U, φ) be a chart for $\mathcal{M}$. The map $\varphi \circ f^{-1}$ is a chart for $\mathcal{N}$.

Proposition 2.41. *Let $f : \mathcal{M} \to \mathcal{N}$ and $g : \mathcal{N} \to \mathcal{K}$ be smooth maps between manifold $\mathcal{M}$, $\mathcal{N}$, and $\mathcal{K}$. Then $g \circ f$ is a smooth map and the chain rule holds:*

$$T(g \circ f) = Tg \circ Tf.$$

Curves on Manifolds   The time evolution of dynamical systems is described by curves on manifolds. We will study such curves next.

Definition 2.69. *Let $\mathcal{M}$ be a manifold. A **curve** $c(t)$ **on** $\mathcal{M}$ is a smooth mapping $c(t) : [a, b] \to \mathcal{M}$. A **curve at m** is a curve such that $a < 0 < b$ and $c(0) = m \in \mathcal{M}$.*

Note that a curve is defined "inverse" to the chart map and it maps from Euclidean space onto a manifold. To define the tangent vector of a curve an appropriate notion of differentiability is needed.

Definition 2.70. *A curve $c(t) : [a, b] \to \mathcal{M}$ is **differentiable** if the **local representative** $(\varphi_i \circ c(t)) : [a, b] \to \mathbf{E}$ in the model space $\mathbf{E}$ of $\mathcal{M}$ is differentiable for all charts (U_i, φ_i) of $\mathcal{M}$ with $U_i \cap c(t) \neq \varnothing$. The derivative $\mathrm{D}(\varphi_i \circ c(t))$ of the local representation of a curve will be denoted by $(\varphi_i \circ c(t))'$.*
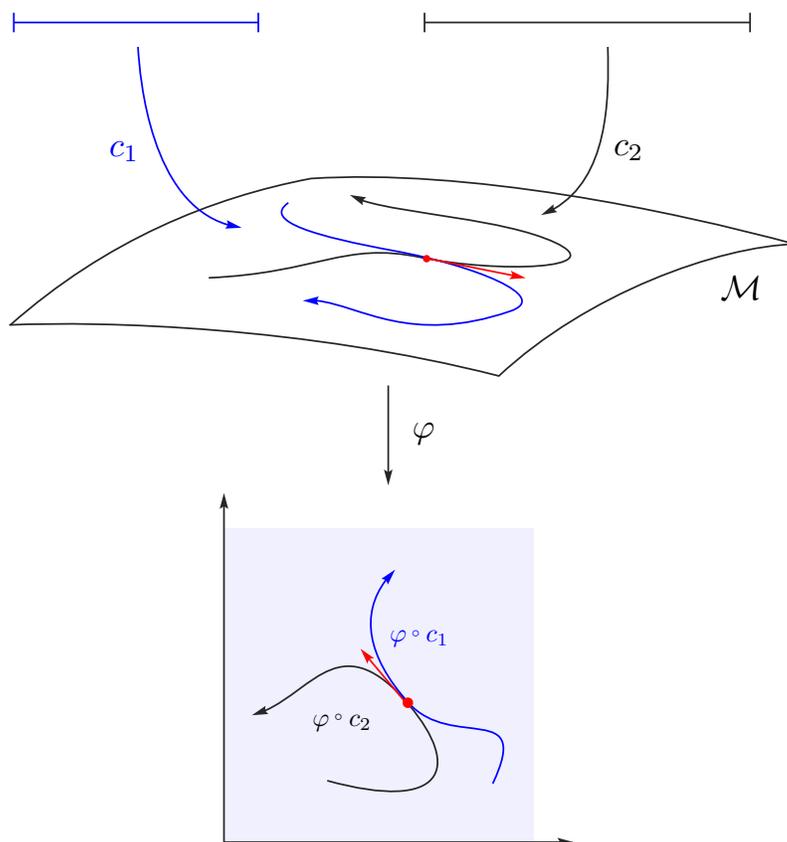
Figure 2.22: The tangent on a manifold $\mathcal{M}$ is defined through curves in the model space.

The local representative $(\varphi_i \circ c(t))$ is a map between normed spaces. Differentiability is hence well defined by Def. 2.56, and for some $\bar{t} \in [a, b]$ the derivative $(\varphi \circ c)'(\bar{t})$ is an element in $\mathbf{E}$. Our notion of differentiability for curves does include the endpoints $\{a, b\}$. This is possible by employing one-sided limits at a and b and it can be shown that these limits are independent of the coordinate chart that is used. A definition including the endpoints is important for manifolds with boundary that we will introduce shortly.

Definition 2.71. *Two curves c_1, c_2 at $m \in \mathcal{M}$ are **tangent at** m if their local representatives satisfy $(\varphi \circ c_1)'(0) = (\varphi \circ c_2)'(0)$ for a chart (U, φ) with $m \in U$. An equivalence class $[c]_m$ of curves tangent at m will be denoted as **tangent vector** X_m.*

The above definition of tangency requires that curves agree in their positions at $t = 0$ and that their tangents vectors $(\varphi \circ c_i)'(0)$ in $\mathbf{E}$ are identical in direction

and magnitude. Considered as maps $c : [a, b] \to \mathbf{E}$, curves are hence tangent if their tangents $Tc : [a, b] \times \mathbb{R} \to \mathbf{E} \times \mathbf{E}$ in the sense of Def. 2.59 agree at $t = 0$, cf. Fig. 2.22. It can be shown that the definition of tangency for a curve, and hence also the concept of a tangent vector X_m, is independent of the chart used. We are now prepared to introduce the important concept of a tangent space.

Definition 2.72. *For a manifold $\mathcal{M}$ the **tangent space** $T_m \mathcal{M}$ **at** $m \in \mathcal{M}$ is the set of all equivalence classes*

$$T_m \mathcal{M} \equiv \{ X_m = [c]_m \mid c \text{ a curve at } m \in \mathcal{M} \} .$$

For a subset $U \subset \mathcal{M}$ let $T\mathcal{M}|_U$ be defined as the disjoint union

$$T\mathcal{M} \mid_U = \bigcup_{m \in U} T_m \mathcal{M}.$$

*The **tangent bundle** of $\mathcal{M}$ is $T\mathcal{M} \mid_{\mathcal{M}}$. The **tangent bundle projection** $\tau_{\mathcal{M}} : T\mathcal{M} \to \mathcal{M}$ of $\mathcal{M}$ is the trivialization $\tau_{\mathcal{M}}([c]_m) = m$.*

Analogous to the derivative of a one dimensional function, the tangent space can be considered as the best linear approximation to a manifold, and it is often this picture that one has in mind, and which agrees with the usual notion of tangent spaces in $\mathbb{R}^3$.

Remark 2.43. A tangent vector is a tuple $(m, X) \in T_m \mathcal{M}$. The base point is often omitted but it is essential for the vector to make sense.

Various alternative definition of the tangent space and the tangent bundle exist. The one presented here is the one most relevant in the context of dynamical systems. Its slightly complicated nature arises from our requirement that geometric objects are intrinsic and defined without reference to an "ambient" space. The tangent space $T_m \mathcal{M}$ at m has a natural vector space structure, as can be shown using its definition based on the linear model space $\mathbf{E}$.

Example 2.36. The tangent space $T_e \mathbf{E}$ of a linear space $\mathbf{E}$ at $e \in \mathbf{E}$ is isomorphic to $\mathbf{E}$. Hence $T\mathbf{E} \cong \mathbf{E} \times \mathbf{E}$.

It can be shown that curves c_1, c_2 that are tangent at $m \in \mathcal{M}$ in the sense of Def. 2.71 are also tangent at $f(m) \in \mathcal{N}$. This suggests the following definition.

Definition 2.73. *Let $f : \mathcal{M} \to \mathcal{N}$ be a map between manifolds. The mapping $Tf([c]_m) = [f \circ c]_{f(m)}$ is the **tangent** or **tangent map** of f.*

With Def. 2.70, a curve $c(t)$ can be identified with its local representative $(\varphi \circ (t)) : [a, b] \to \mathbf{E}$ in a chart Def. 2.73 is hence a special case of the tangent map between linear spaces defined in Def. 2.59, that also provides a concrete expression for $Tf([c]_m)$. The above tangent of a mapping is fiber-preserving in that a curve tangent at m will be mapped to a curve tangent at $f(m)$. This motivates the concept of the derivative of a map that relates the tangent spaces of two manifolds, cf. Fig. 2.23.
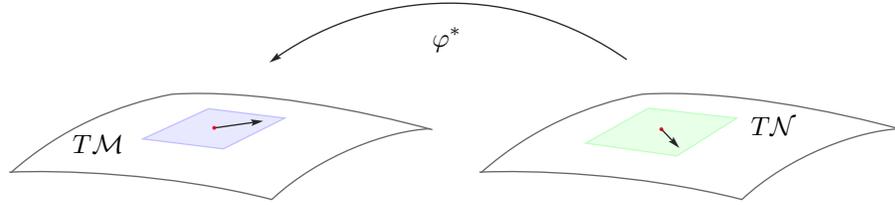
Figure 2.23: The derivative $T_m f$ of a map takes the tangent space $T_m \mathcal{M}$ at $m \in \mathcal{M}$ to the tangent space $T_{f(m)}\mathcal{N}$ at $f(m) \in \mathcal{N}$.

Definition 2.74. *The **derivative** $T_m f : T_m \mathcal{M} \to T_{f(m)}\mathcal{N}$ of a smooth map $f : \mathcal{M} \to \mathcal{N}$ at $m \in \mathcal{M}$ is the restriction $Tf \mid_{T_m \mathcal{M}}$ of Tf to the tangent space $T_m \mathcal{M}$ at $m \in \mathcal{M}$.*

Since $T_m \mathcal{M}$ and $T_{f(m)}\mathcal{N}$ are linear spaces also the map $T_m f$ is linear for fixed $m \in \mathcal{M}$. If $T_m f : T_m \mathcal{M} \to T_{f(m)}\mathcal{N}$ is an isomorphism then the inverse function theorem in Theorem 2.11 asserts that f is a local diffeomorphism, that is there exist open neighborhoods $U(m) \in \mathcal{M}$ and $V(f(m)) \in \mathcal{N}$ such that $f \mid_U : U \to V$ is a diffeomorphism.

Remark 2.44. Let $\mathcal{M} \subset \mathbb{R}^m$ be an n-dimensional embedded manifold as in Remark 2.41. The inverse chart map $\varphi^{-1} : \mathbb{R}^n \to \mathbb{R}^m$ with coordinates $\varphi^{-1}(x_1, \ldots x_n) = (y^1(x^1, \ldots, x^n), \ldots, y^m(x^1, \ldots, x^n))$ then relates the model space $\mathbb{R}^n$ to the embedding space $\mathbb{R}^m$, and the tangent $T\varphi^{-1}$ maps local representatives $(\varphi \circ c)(t)$ in the chart $\mathbb{R}^n$ to curves in $\mathbb{R}^m$. On $\mathbb{R}^m$, the derivative is well defined, recovering the usual notion of a tangent vector "attached" to a manifold. For an embedded manifold, $T\varphi^{-1}$ hence relates tangent vectors in the model space $\mathbb{R}^n$ to tangent vectors on $\mathcal{M}$ in the embedding space $\mathbb{R}^m$. Using the expression for the tangent between linear spaces in Def. 2.59, $(Tf)(u, e) = (f(u), (Df) \cdot e)$, and using Remark 2.40 the tangent of the inverse chart map is given by by the Jacobian matrix

$$T_x \varphi^{-1} = J(x) = \begin{bmatrix} \partial y_1^i / \partial x_1 & \cdots & \partial y_1^i / \partial x_n \\ \vdots & \ddots & \vdots \\ \partial y_n^i / \partial x_1 & \cdots & \partial y_n^i / \partial x_n \end{bmatrix} \tag{2.75}$$

which maps tangent vectors in the chart $\mathbb{R}^n$ to those in the embedding space $\mathbb{R}^m$.

Remark 2.45. Reasoning similar to the above remark shows that also the local expression for the tangent $T\varphi_{ij}$ of the transition map $\varphi_{ij} = \varphi_i \circ \varphi_j^{-1}$ between two charts (U_i, φ_i) and (U_j, φ_j) is given by a Jacobian matrix.

Remark 2.46. For a finite dimensional manifold $\mathcal{M}$, the tangent map $T_{\varphi(m)}\varphi^{-1} : T_{\varphi(m)}\mathbb{R}^n \to T_m \mathcal{M}$ can be employed to map iso-coordinate curves from the model space $\mathbb{R}^n$ onto $\mathcal{M}$, analogous to how latitude and longitude lines are rendered
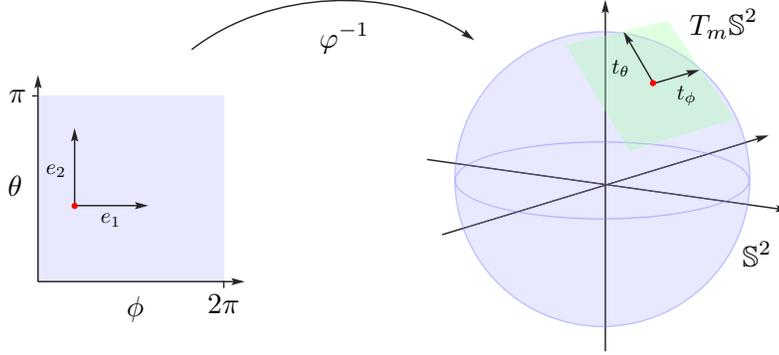
Figure 2.24: Tangent space $T_m S^2$ of the sphere S^2 at m with basis vectors t_θ and t_ϕ that are obtained from the canonical basis vectors e_1, e_2 in the chart.

on geographic maps. When $\mathcal{M}$ is embedded in $\mathbb{R}^m$, a basis for the tangent space $T_m \mathcal{M}$ is then given by the derivatives of these curves in the embedding space. By Remark 2.44, the basis vectors are given by

$$\left(T_x\, \varphi^{-1} \right) e_i = J(x)\, e_i \tag{2.76}$$

where $\{e_1, \ldots, e_n\}$ is the canonical basis for $\mathbb{R}^n$. The ith column of the Jacobian $J(x)$ thus provides the ith tangent vector t_i which by the definition of the Jacobian can be written as

$$t_i = \left(\frac{\partial \varphi_1^{-1}}{\partial x^i}, \cdots, \frac{\partial \varphi_m^{-1}}{\partial x^i} \right) = \frac{\partial}{\partial x^i} \left(\varphi_1^{-1}, \cdots, \varphi_m^{-1} \right). \tag{2.77}$$

To emphasize that the basis is independent of the specific chart, and with slight abuse of notation, one often writes $\partial/\partial x^i$ for the ith basis vector, and the notation $\partial/\partial x^i$ is commonly used for arbitrary finite dimensional manifolds, embedded or not, see also Remark 2.56.

Example 2.37. Continuing Example 2.35 and with Remark 2.44, the tangent space $T_m S^2$ of the sphere $S^2 \subset \mathbb{R}^3$ is spanned by vectors $(t_\theta, t_\phi) \in \mathbb{R}^3$ given by

$$t_\theta = \frac{\partial \varphi^{-1}}{\partial \theta} = \begin{pmatrix} -\sin\theta \sin\phi \\ \sin\theta \cos\phi \\ 0 \end{pmatrix} \qquad t_\phi = \frac{\partial \varphi^{-1}}{\partial \phi} = \begin{pmatrix} -\cos\theta \cos\phi \\ \cos\theta \sin\phi \\ -\sin\theta \end{pmatrix}, \tag{2.78}$$

cf. Fig. 2.24. The vectors are the images of the canonical basis vectors $e_1 = (1,0)^T$ and $e_2 = (0,1)^T$ in the model space $\mathbb{R}^2$ under the tangent map $T\varphi^{-1}$, that is

$$t_\theta \frac{\partial \varphi^{-1}}{\partial \theta} = J\, e_1 \qquad t_\phi \frac{\partial \varphi^{-1}}{\partial \phi} = J\, e_2 \tag{2.79}$$
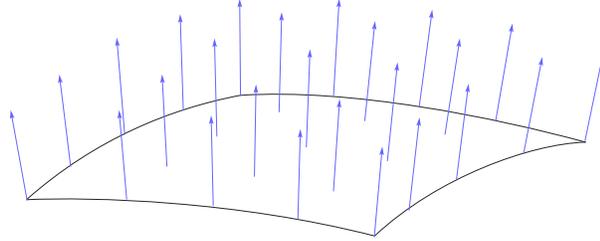
Figure 2.25: Line bundle (blue) over a manifold $\mathcal{M}$. At every point of $\mathcal{M}$ a vector space, $\mathbb{R}$, is "attached".

where J is the Jacobian in Eq. 2.75 for the coordinate expression for spherical coordinates in Eq. 2.74b given by

$$J = \begin{pmatrix} \partial\varphi_x^{-1}/\partial\theta & \partial\varphi_x^{-1}/\partial\phi \\ \partial\varphi_y^{-1}/\partial\theta & \partial\varphi_y^{-1}/\partial\phi \\ \partial\varphi_z^{-1}/\partial\theta & \partial\varphi_z^{-1}/\partial\phi \end{pmatrix} = \begin{pmatrix} -\sin\theta\sin\phi & -\cos\theta\cos\phi \\ \sin\theta\cos\phi & \cos\theta\sin\phi \\ 0 & -\sin\theta \end{pmatrix} \tag{2.80}$$

and it is easy to see that this agrees with Eq. 2.78.

It can be shown that the tangent bundle $T\mathcal{M}$ is a manifold in its own right, $2n$ dimensional if $\mathcal{M}$ is n dimensional, and that an atlas for $T\mathcal{M}$ is induced by an atlas for $\mathcal{M}$. This atlas is sometimes known as natural atlas for $T\mathcal{M}$.

Remark 2.47. Since $T\mathcal{M}$ is a manifold one can consider "higher order" tangent spaces such as $T(T\mathcal{M}) = TT\mathcal{M}$, and the resulting hierarchy of manifolds is sometimes known as a tower. In the context of such constructions it is important to distinguish vector field *on* a manifold $T\mathcal{M}$, which lie in $T(T\mathcal{M})$, and vector fields *in* $T\mathcal{M}$, which lie in $T\mathcal{M}$.

Remark 2.48. It is important to note that the tangent bundle $T\mathcal{M}$ does in general *not* have the structure of the tensor product $T \times \mathcal{M}$. The tangent bundle TS^1 of the circle S^1 is indeed diffeomorphic to $\mathbb{R} \times S^1$, but already for the 2-sphere this is no longer true. There the obstruction to a trivial tangent bundle is closely related to the fact that no global non-vanishing vector field on the sphere exists.

A vector bundle is the generalization of the tangent bundle where an arbitrary vector space is "attached" to each base point $m \in \mathcal{M}$. For example instead of "attaching" a plane to each point of S^2, which is the blueprint for the tangent space TS^2, one can "attach" a line oriented along the normal direction, which would be similar to the Gauss map, see Fig. 2.25. To formally introduce the concept of a vector bundle we will first define the linear analogue, a local vector bundle, and then extend the concept to the nonlinear setting, a recipe that we will use more frequently later on.

Definition 2.75. *Let* **E** *and* **F** *be Banach space and let* $U \subset \mathbf{E}$ *be open. The tensor product* $U \times \mathbf{F}$ *is a* ***local vector bundle***. U *is the* ***base*** *of* $U \times \mathbf{F}$ *and* $\{u\} \times \mathbf{F}$ *is the* ***fiber*** *over* $u \in U$, *endowed with the vector space structure of* **F**. *The* ***fiber projection*** $\pi : U \times \mathbf{F}$ *is the trivialization* $\pi(u, f) = u$, *and the* ***fiber*** *over* $u \in U$ *is given by* $\pi^{-1}(u)$.

A vector bundle is a manifold, in the sense of Def. 2.65, that looks locally like a local vector bundle. To smoothly "stitch together" multiple local vector bundles we need local vector bundle maps.

Definition 2.76. *Let* $\mathbf{E}, \mathbf{F}$ *and* $\mathbf{E}', \mathbf{F}'$ *be Banach spaces and* $U \subset \mathbf{E}$ *and* $U' \subset \mathbf{E}'$ *be open. A* ***local vector bundle map*** $\varphi : U \times \mathbf{F} \to U' \times \mathbf{F}'$ *is a smooth map such that for* $u \in U$ *and* $f \in \mathbf{F}$ *the image is given by* $\varphi(u, f) = (\varphi_B(u), \varphi_F(f))$, *where* $\varphi_B : U \to U'$ *and* $\varphi_F : U \in L(\mathbf{F}, \mathbf{F}')$ *are smooth. A local vector bundle map that has an inverse that is also a local vector bundle map is a* ***vector bundle isomorphism***.

Local vector bundle isomorphisms provide the transitions maps between the charts of a vector bundle, which we are now prepared to introduce.

Definition 2.77. *Let* S *be a set. A* ***local bundle chart*** (W, φ), *a subset* $W \subset S$ *together with a bijective vector bundle chart map* $\varphi : W \to U \times \mathbf{F}$, *gives* S *locally the structure of a local vector bundle* $U \times \mathbf{F}$, *which may depend on* φ. *A* ***vector bundle*** E *is a manifold* $(S, \mathcal{B})$ *where* $\mathcal{B}$ *is a differentiable vector bundle structure for* S *given by an equivalence class of vector bundles atlases formed by local bundle charts whose transition maps are local vector bundle isomorphisms.*

As for manifolds, a vector bundle $(S, \mathcal{B})$ will usually be identified with the underlying set S and we will write E when we refer to the tuple $(S, \mathcal{B})$.

Definition 2.78. *Let* $E = (S, \mathcal{B})$ *be a vector bundle. The* ***base*** B *of* E *is the union of the base of all local vector bundles* $U \times \mathbf{F}$,

$$B = \left\{ e \in E \mid \exists (W, \varphi) : e = \varphi^{-1}(u, 0) \right\}.$$

The ***fiber projection*** *of* **E** *is the map* $\pi : E \to B$ *given by* $\pi(b, f) = b$, *and* $\pi^{-1}(b)$ *is the* ***fiber*** *over* $b \in B$. *A vector bundle is* ***trivial*** *when it has the structure of a tensor product.*

In the above definition it is critical that the vector bundle chart map φ is a bijection for the pre-image to correspond to the intuitive notion of the base. It can be shown that the base B is a submanifold of the vector bundle E and that the bundle projection $\pi : E \to B$ is smooth and surjective. The fiber $\pi^{-1}(b)$ for $b \in B$ has an intrinsic vector space structure with b being the zero element. In the following, a vector bundle will often be identified with the bundle projection $\pi : E \to B$.

Example 2.38. The tangent bundle $T\mathcal{M}$ is a vector bundle $\tau_{\mathcal{M}} : T\mathcal{M} \to \mathcal{M}$ whose fibers $T_m\mathcal{M}$ are isomorphic to the model space of $\mathcal{M}$.

Example 2.39. The cylinder $C = S^1 \times \mathbb{R}$ is a vector bundle with $\pi : S^1 \times \mathbb{R} \to S^1$. It is a trivial vector bundle since it is a tensor product.

Vector bundle maps can be defined analogously to tangent maps, cf. Def. 2.73. In particular, a vector bundle map $f : E \to \bar{E}$ preserve the base, in that $f(B) \subset \bar{B}$, and the fibers, in that $f_B \circ \pi_E = \pi_{\bar{E}} \circ f$ for the restriction f_B of f to the base, and the following diagram commutes

$$
\begin{array}{ccc}
E & \xrightarrow{\ f\ } & \bar{E} \\
\pi_E \downarrow & & \downarrow \pi_{\bar{E}} \\
B & \xrightarrow{\ f_B\ } & \bar{B}
\end{array}
\tag{2.81}
$$

When $\mathbb{R}^2$ is considered as the trivial vector bundle $\pi = \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ then a smooth function $f : \mathbb{R} \to \mathbb{R}$ assigns to each abscissa value in the base an ordinate value in the fiber, and values in the fiber vary smoothly for changing base point. This construction can be generalized to arbitrary vector bundles.

Definition 2.79. *Let $\pi : E \to B$ be a vector bundle. A **local vector bundle section** $\xi : U \to E$ for an open set $U \subset B$ is a map such that $\pi(\xi(b)) = b$ for all $b \in U$. A **global vector bundle section** is a local section where $U = B$. The space of all local sections of a vector bundle is denoted by $\Gamma(E)$.*

Evidently, $\xi(b)$ is a map from the base B into the fiber $\pi^{-1}(b)$ over $b \in B$. The space of all section $\Gamma(E)$ has a natural vector space structure analogous to the usual vector space structure for functions.

Remark 2.49. Let $\mathcal{M}$ be a manifold. A function $f : \mathcal{M} \to \mathbb{R}$ is a section of the line bundle over $\mathcal{M}$. The space of all such sections is denoted by $\mathcal{F}(\mathcal{M})$ and it has the structure of a ring with addition and multiplication defined as usual in each fiber.

Fiber bundles are generalizations of vector bundles where instead of a vector space an arbitrary manifold is "attached" to each point of the base space. Most of the definitions and concepts for vector bundles carry over with obvious modifications and we will therefore omit their formal definition. An example of a fiber bundle is the sphere bundle $S^2\mathcal{M}$ over a two manifold $\mathcal{M}$ in $\mathbb{R}^3$.

Submanifolds   Submersions and immersions are important means to obtain submanifolds with well defined properties.

Definition 2.80. *Let $\mathcal{M}, \mathcal{N}$ be manifolds and $f : \mathcal{M} \to \mathcal{N}$ be a smooth map. A point $m \in \mathcal{M}$ is a **regular point** of f is $T_m f$ is surjective; otherwise m is a **critical point**. Correspondingly, the image $f(m) \in \mathcal{N}$ of a regular point is a **regular value** and the image of a critical point is a **critical value**. The mapping f is a **submersion** if it does not have critical points.*

Sard's theorem states that, under suitable smoothness assumptions, the set of regular values of f is dense in $\mathcal{N}$. Submersions are open mappings, that is they map open sets to open sets. The importance of submersion largely stems from the submersion theorem.

Theorem 2.13 (Submersion Theorem). *Let $f : \mathcal{M} \to \mathcal{N}$ be a smooth mapping between manifolds and $n \in \mathcal{N}$ be a regular value of f. Then the pre-image*

$$\mathcal{K}(n) = \left\{ m \in \mathcal{M} \mid m = f^{-1}(n) \, , \, n \in \mathcal{N} \right\}$$

is a closed submanifold of $\mathcal{M}$ with tangent space $T_m \mathcal{K}(n) = \ker(T_m f)$. If $\mathcal{N}$ is finite dimensional then $\operatorname{codim}(\mathcal{K}) = \dim(\mathcal{N})$.

Definition 2.81. *A smooth map $f : \mathcal{M} \to \mathcal{N}$ between manifolds is an **immersion** at $m \in \mathcal{M}$ if $T_m f$ is injective. A global immersion is an immersion for every $m \in \mathcal{M}$ and $f(\mathcal{M})$ is then an **immersed submanifold** of $\mathcal{N}$.*

Analogous to submersions, the importance of immersions arises from an associated theorem.

Theorem 2.14 (Immersion Theorem). *A smooth map $f : \mathcal{M} \to \mathcal{N}$ is an immersion at $m \in \mathcal{M}$ if and only if there is a neighborhood U of m such that $f(U)$ is a submanifold in $\mathcal{N}$ and $f\mid_U$ is a local diffeomorphism between U and $f(U)$.*

Intuitively, the difference between submersions and immersions can be related to the dimensionality of $\mathcal{M}$ and $\mathcal{N}$. If $\dim(\mathcal{M}) > \dim(\mathcal{N})$ then f cannot be an immersion and it is a submersion if the rank of the tangent map equals the dimension of $\mathcal{N}$ everywhere. Conversely, if $\dim(\mathcal{M}) < \dim(\mathcal{N})$ then f cannot be a submersion and it is an immersion if it provides a local diffeomorphism between $\mathcal{M}$ and $\mathcal{N}$. The image $f(\mathcal{M})$ of an immersion does not have to be a submanifold and the topologies of $f(\mathcal{M})$ and $f(\mathcal{N})$ do not have to agree. This does not contradict the immersion theorem since locally $f(U)$ can be a submanifold but there might be global obstructions.[101] A stronger notion guarantees the submanifold property.

Definition 2.82. *An immersion $f : \mathcal{M} \to \mathcal{N}$ is an **embedding** if f is a homeomorphism onto $f(\mathcal{M})$ with respect to the topology of $\mathcal{N}$.*

It immediately follows that if $f : \mathcal{M} \to \mathcal{N}$ is an embedding then $f(\mathcal{M})$ is a submanifold of $\mathcal{N}$. Another important notion in the context of submanifolds is transversality.

Definition 2.83. *A smooth map $f : \mathcal{M} \to \mathcal{N}$ is **transversal** with respect to a submanifold $\mathcal{K} \subset \mathcal{N}$, denoted by $f \pitchfork \mathcal{K}$, if $f^{-1}(\mathcal{K}) = \emptyset$ or $(T_m f)(T_m \mathcal{M}) + T_{f(m)} \mathcal{K} = T_{f(m)} \mathcal{N}$ and the inverse image $(T_{f(m)} f^{-1})(T_{f(m)} \mathcal{K})$ splits in $T_m \mathcal{M}$ for every $f^{-1}(\mathcal{K}) = m \in \mathcal{M}$.*

---

[101]See (Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*, p. 177) for an example.

Note that the second condition on the inverse image is automatically satisfied if $\mathcal{M}$ is finite dimensional or Hilbert; and in the first condition no assumption on the vector space sum is made and it is not required that the spaces provide an orthogonal decomposition of $T_{f(m)}\mathcal{N}$.

Manifolds with Boundary Many manifolds in practice do not satisfy Def. 2.65. For example, the three-ball $\mathcal{B}^3 = \{\|x\| \leq 1\}$ in $\mathbb{R}^3$ is not a manifold since the points with $\|x\| = 1$ do not have a neighborhood homeomorphic to $\mathbb{R}^3$. To extend our concept of a manifold to these situations manifolds with boundary are needed. As usual, we will define the necessary concepts first for the model space $\mathbf{E}$ and then "transfer" them to the manifold.

Definition 2.84. *Let $\mathbf{E}$ be a Banach space and $\mathbf{E}^*$ its dual. The **half space** $\mathbf{E}_\lambda$ with respect to $\lambda \in \mathbf{E}^*$ is*

$$\mathbf{E}_\lambda = \{e \in \mathbf{E} \mid \lambda(e) \geq 0\}.$$

*For $U \subset \mathbf{E}_\lambda$ an open set, the **interior** $\mathrm{int}\,(U)$ of U is*

$$\mathrm{int}\,(U) = U \cap \{e \in E \mid \lambda(e) > 0$$

*and the **boundary** ∂U of U is*

$$\partial U = U \cap \ker(\lambda).$$

It follows that $U = \mathrm{int}\,(U) \cup \partial U$ and $\mathrm{int}\,(U) \cap \partial U = \emptyset$, and $\mathrm{int}\,(U)$ and ∂U form a cover for U. To define a manifold with boundary based on a half space, we need to define what it means for transition maps to be smooth at the boundary.

Definition 2.85. *Let $\mathbf{E}$ and $\mathbf{F}$ be Banach spaces with dual space $\mathbf{E}^*$ and $\mathbf{F}^*$, $\lambda \in \mathbf{E}^*$ and $\mu \in \mathbf{F}^*$, and $U \subset \mathbf{E}_\lambda$ and $V \subset \mathbf{F}_\mu$ be open. A map $f : U \to V$ is **smooth** if for every $x \in U$ there are open neighborhoods $U(x)$ and $V(f(x))$ and a smooth map $f_x : U(x) \to V(f(x))$ such that $f\mid_{U \cap U(x)} = f_x\mid_{U \cap U(x)}$ and $\mathrm{D}f(x) = \mathrm{D}f_x(x)$. The map f is a **diffeomorphism** if there is a smooth map $g : V \to U$ which is an inverse of f.*

It can be shown that the definition of $\mathrm{D}f$ is independent of the choice of f_x.

Definition 2.86. *Let S be a set. A **chart with boundary** (U, φ) is a set $U \subset S$ and a map $\varphi : U \to V \subset \mathbf{E}_\lambda$ for some $\lambda \in \mathbf{E}^*$. An **atlas of charts with boundary** for S is a family of charts in the sense of Def. 2.63 and smoothness of the transition maps as defined in Def. 2.85.*

After introducing charts with boundary, we are ready to define a manifold with boundary, see Fig. 2.26.
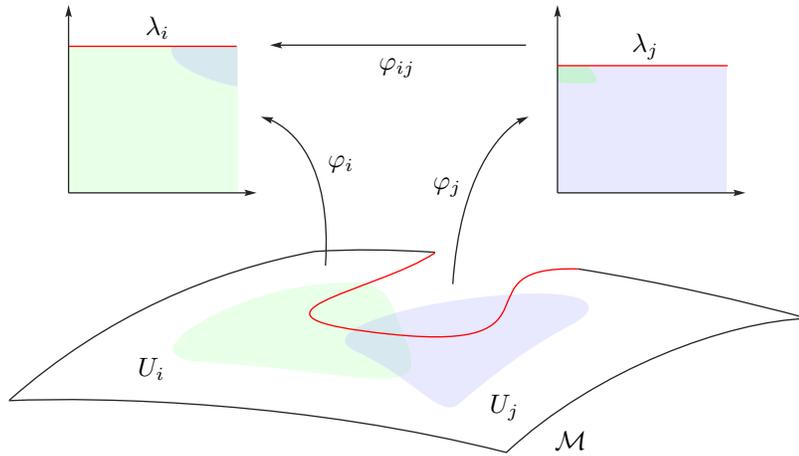
Figure 2.26: A manifold $\mathcal{M}$ with boundary $\partial\mathcal{M}$ (red) with charts (U_i, φ_i) and (U_j, φ_j) that are open half spaces defined through λ_i and λ_j. At the overlap the charts are required to fit together so that the transition map $\varphi_{ij} = \varphi_i \circ \varphi_j^{-1}$ is smooth.

Definition 2.87. *A **manifold with boundary** is a set $\mathcal{M}$ with an atlas of chart with boundary. The **interior** $\mathrm{int}\,(\mathcal{M})$ of $\mathcal{M}$ is*

$$\mathrm{int}\,(\mathcal{M}) = \bigcup_i \varphi_i^{-1}\left(\mathrm{int}\,(\varphi(U_i))\right)$$

*and the **boundary** $\partial\mathcal{M}$ is*

$$\partial\mathcal{M} = \bigcup \varphi_i^{-1}\left(\partial\,(\varphi(U_i))\right).$$

The atlas for a manifold with boundary $\mathcal{M}$ induces an atlas for its interior $\mathrm{int}\,(\mathcal{M})$ and an atlas for its boundary $\partial\mathcal{M}$ by suitable restrictions of the image of the chart map in the model space.

Proposition 2.42. *Let $\mathcal{M}$ be a manifold with boundary. Then its interior $\mathrm{int}\,(\mathcal{M})$ and its boundary $\partial\mathcal{M}$ are manifolds without boundary. Moreover, when $\mathcal{M}$ is finite with dimension n then the $\dim(\mathrm{int}\,(\mathcal{M})) = n$ and $\dim(\partial\mathcal{M}) = n-1$.*

The boundary $\partial\mathcal{M}$ is a manifold without boundary but it might have multiple disconnected components, such as the boundary of a finite cylinder $\mathcal{C}$ that is $\partial\mathcal{C} = \{S^1, S^1\}$.

Proposition 2.43. *Let $f : \mathcal{M} \to \mathcal{N}$ be a diffeomorphism between smooth manifolds $\mathcal{M}, \mathcal{N}$ with boundary. When suitably restricted, f then also provides diffeomorphisms*

$$\mathrm{int}\,(f) \equiv f|_{\mathrm{int}\,(\mathcal{M})} \colon \mathrm{int}\,(\mathcal{M}) \to \mathrm{int}\,(\mathcal{N}) \qquad \partial f \equiv f|_{\partial\mathcal{M}} \colon \partial\mathcal{M} \to \partial\mathcal{N}.$$
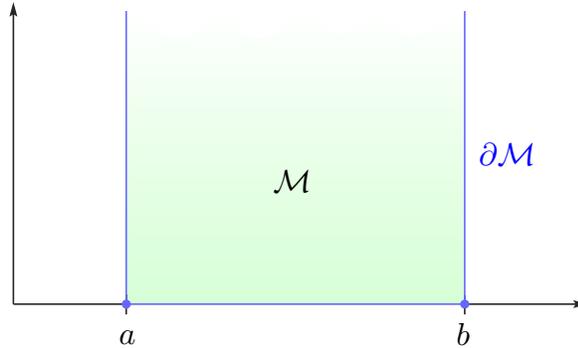
Figure 2.27: $\mathcal{M} = [a, b] \times \mathbb{R}$ is a manifold with boundary $\partial\mathcal{M} = \{a, b\} \times \mathbb{R}$. The boundary $\partial\mathcal{M}$ consists of two manifolds with boundary $a \times \mathbb{R}$ and $b \times \mathbb{R}$, and it is non-compact in the fibers.

The proposition also applies when $\mathcal{M} = \mathcal{N}$, which is of importance in many physical applications. Since closed intervals $[a, b]$ were used to define curves on manifolds, the tangent space of a manifold with boundary is defined as in Def. 2.72 for manifolds without boundary *without* modifications. In particular, in the finite dimensional case $T_m\mathcal{M}$ is isomorphic to the model space of $\mathcal{M}$ even on the boundary $\partial\mathcal{M}$.

Remark 2.50. The tensor product $\mathcal{M} \times \mathcal{N}$ of two manifolds with boundary $\mathcal{M}$ and $\mathcal{N}$ is *not* a manifold with boundary, that is there is no tensor product in the category of manifolds with boundary. An intuitive way to see this is to consider the boundaries $\partial\mathcal{M}$ and $\partial\mathcal{N}$ that are $(n-1)$- and $(k-1)$-dimensional manifolds when $\mathcal{M}$ and $\mathcal{N}$ are n- and k-dimensional, respectively. But then $\partial\mathcal{M} \times \partial\mathcal{N}$ is $(n+k-2)$-dimensional while the boundary $\partial(\mathcal{M} \times \mathcal{N})$ is required to be $(n+k-1)$-dimensional. For a manifold *with* boundary $\mathcal{M}$ and a manifold *without* boundary $\mathcal{N}$ the product manifold $\mathcal{M} \times \mathcal{N}$ is a manifold with boundary and $\partial(\mathcal{M} \times \mathcal{N}) = \partial\mathcal{M} \times \mathcal{N}$, see Fig. 2.27.[102]

Vector Fields and Flows   Dynamical systems are generated by vector fields and physically observable quantities are transported along their flow. Vector fields are hence central to geometric mechanics and they will be introduced next.

Definition 2.88. *A **vector field** $X(\mathcal{M})$ on a manifold $\mathcal{M}$ is a section $X \in \Gamma(T\mathcal{M})$ of the tangent bundle $\mathcal{M}$. The **support** of $X(\mathcal{M})$ is the set $\{m \in \mathcal{M} \mid X(m) \neq 0\}$. The space of all vector fields on $\mathcal{M}$ is denoted by $\mathfrak{X}(\mathcal{M})$.*

A vector field $X : \mathcal{M} \to T\mathcal{M}$ assigns to every tangent space a vector $X(m) \in T_m\mathcal{M}$, and the vectors are changing smoothly as a function of the surface location. A time-dependent vector field $X(m, t)$ is defined analogously as $X_t : \mathcal{M} \times \mathbb{R} \to T\mathcal{M}$.

[102]See for example (Outerelo and Ruiz, *Mapping degree theory*, p. 57).
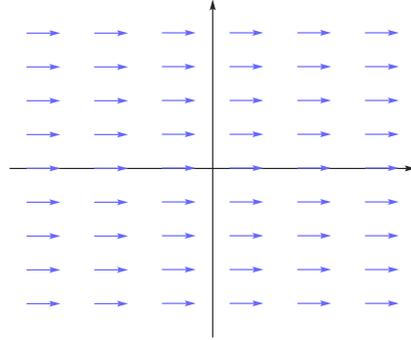
Figure 2.28: Vector field associated with the canonical basis vector e_1 when considered for each tangent space $T_x \mathbb{R}^2$.

Example 2.40. Let $\mathcal{M} = \mathbb{R}^2$. The canonical coordinate axis e_1 for $T_x \mathbb{R}^2 \cong \mathbb{R}^2 \times \mathbb{R}^2$, considered for all $x \in \mathbb{R}^2$, defines a vector fields on $\mathbb{R}^2$, cf. Fig. 2.28

Example 2.41. The negative gradient $-\nabla f$ of the potential $f(x) = \|x\|$, $x \in \mathbb{R}^3$, defines a vector field on $\mathbb{R}^3 \setminus \{0\}$, cf. Def. 2.60.

Definition 2.89. *Let $\mathcal{M}$ be a manifold. The **local representative** of a vector field $X \in \mathfrak{X}(\mathcal{M})$ for a chart (U, φ) is the vector field*

$$X_{\mathrm{lr}} = (T\varphi)(X|_U) \in T\mathbf{E} \cong \mathbf{E} \times \mathbf{E},$$

where $T\varphi$ is the tangent of the chart map φ.

A vector field and its local representative are usually identified, which is possible since even with the local representative the result is independent of the chart used. However, for conceptual questions it can be important to keep the difference in mind.

Remark 2.51. When the model space for the manifold $\mathcal{M}$ is $\mathbb{R}^n$, the local representative of a vector field $X \in \mathfrak{X}(\mathcal{M})$ is given by an n component vector $(X^1, \ldots, X^n)$ with respect to the canonical basis $\{e_1, \ldots, e_n\}$ at every $x \in \mathbb{R}^n$,

$$X_{\mathrm{lr}}(x) = X^1 e_1 + \ldots + X^n e_n. \tag{2.82}$$

Smoothness then means that the components $X^i(x)$ vary smoothly as a function of the location $x \in \mathbb{R}^n$ in the chart.

Remark 2.52. For reasons that will become clear in the sequel the correct usage of "upstairs" and "downstairs" indices is critical for vectors and their generalization, cf. Eq. 2.82.

Definition 2.90. *Let $X(\mathcal{M})$ be a vector field on an manifold $\mathcal{M}$. An **integral curve of** X **at** $m \in \mathcal{M}$ is a curve $c(t) : [a, b] \to \mathcal{M}$ at m such that $c'(t) = X(c(t))$ for all $t \in [a, b]$. For a time dependent vector field an integral curve satisfies $c'(t) = X(c(t), t)$.*

126

Using the chain rule, it is easy to show that a vector field is an integral curve if and only if $(\varphi \circ c)'(t) = X_{\text{lr}}((\varphi \circ c)(t))$. The condition $c'(t) = X(c(t))$ in Def. 2.90 is the differential equation $Dc(t) = X(c(t))$. In local coordinates it takes the form

$$\frac{dc^1(t)}{dt} = X^1(c^1(t), \ldots, c^n(t))$$

$$\vdots \qquad\qquad \vdots$$

$$\frac{dc^n(t)}{dt} = X^n(c^1(t), \ldots, c^n(t)).$$

Hence, questions about existence and uniqueness of solutions $c(t)$ arise. Solutions to systems of ordinary differential equations are a well studied subject, and using the local representatives known results immediately apply. In particular, it can be shown that for a smooth vector field integral curves exist and these are locally unique and smooth. Globally, only uniqueness can be proven. Analogous results can be shown for time dependent vector fields.

Definition 2.91. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$ a smooth vector field. A **local flow of** X **at** $m \in \mathcal{M}$ is the triple (U, a, F) with $m \in U$ satisfying*

i) The flow map $F_t : U \times I_a \to U_t \subset \mathcal{M}$ is a smooth map for $I_a =] - a, a[$,

ii) $c_u(t) = F(u, t)$ is an integral curve of X at u for all $u \in U$,

iii) $F_t : U \to \mathcal{M}$ is a diffeomorphism onto its image $F_t(U) \subset \mathcal{M}$ for all $t \in I_a$ and $U_t = F_t(U)$ is open.

*The mapping F_t is also known as **evolution operator** since it describes the time evolution of the set U.*

Intuitively, consider $u \in U$ and the integral curve $c_u(t)$ at u. The disjoint union of all such curves is then the flow map F_t, see Fig. 2.29. A local flow is sometimes also referred to as a flow box.

Example 2.42. Consider the vector field in Example 2.40, cf. Fig. 2.28 For any compact $U_0 \subset \mathbb{R}^2$ the flow map is given by $F_t : U_0 \to U_0 + te_1$.

Proposition 2.44. *For $t, s, t + s \in I_a$ the local flow satisfies the **group property***

$$F_{t+s} = F_t \circ F_s = F_s \circ F_t \tag{2.83}$$

with F_0 being the identity map I and $F_{-t} \circ F_t = I$.

Next to local properties of the flow generated by a vector field also global ones that extend as far as possible in time are of interest.
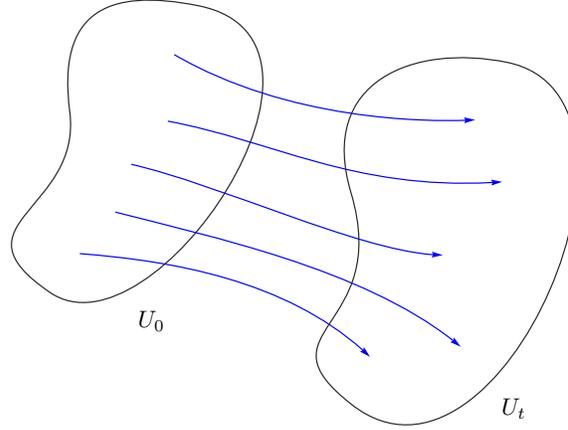
Figure 2.29: Local flow mapping U_0 to U_t by the flow map F_t that for every $u \in U$ is defined through the integral curve $c_u(t)$ at u (blue).

Definition 2.92. *Let $\mathcal{M}$ be a manifold and $X(\mathcal{M}) \in \mathfrak{X}(\mathcal{M})$ a vector field. The **flow domain** $\mathcal{D}_X \subset \mathcal{M} \times \mathbb{R}$ of X is defined as all (m, t) such that there exists an integral curve $c_m : I_m \times \mathcal{M} \to \mathcal{M}$ of X at $m \in \mathcal{M}$ for $t \in I_m$. A vector field X is called **complete** if $\mathcal{D}_X = \mathcal{M} \times \mathbb{R}$ and it is σ-**complete**, with $\sigma \in \{+, -, \pm\}$, when $\mathcal{D}_X \cap \{m\} \times \mathbb{R}$ contains all (m, t) for $t > 0$, $t < 0$, and $t \in \mathbb{R}$, respectively. The positive and negative **lifetime** of m with respect to X are $T^+(m) = \sup(I_m)$ and $T^-(m) = \inf(I_m)$, respectively.*

Example 2.43. The vector field in Example 2.40 is complete.

Example 2.44. The vector field in Example 2.40 defined over the punctured plane $\mathbb{R}^2 \setminus \{0\}$ is not complete.

Intuitively, $\mathcal{D}_X$ assigns to every m the possible time I_m of integral curves. For a complete flow $T^+(m) = \infty$ and $T^-(m) = -\infty$. Completeness corresponds to dynamics that are well defined for all times. In reality, this is never satisfied, for example due to the formation of shocks, but it is important in many idealized settings such as the ideal Euler fluid.

Definition 2.93. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(M)$. The map $F_X : \mathcal{D}_X \to \mathcal{M}$ associated with X such that $t \to F_X(m, t)$ is an integral curve at m is the **integral** of X, and the integral curve $t \to F_X(\mathcal{M}, t)$ is the **maximal integral curve** of X. If X is complete then F_X is the **flow** of X.*

The following proposition summarizes the properties of global flows and ensures the existence of integrals of vector fields.

Proposition 2.45. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$. Then*

i) $\mathcal{D}_X \subset \mathcal{M} \times 0$;

ii) $\mathcal{D}_X$ is open in $\mathcal{M} \times \mathbb{R}$;

iii) there is a unique integral F_X for all $m \in \mathcal{M}$;

iv) for $(m,t) \in \mathcal{D}_X$ the point $(F_X(m,t),s) \in \mathcal{D}_X$ if and only if $(m,t+s) \in \mathcal{D}_X$ and then $F_X(m,t+s) = F_X(F_X(m,t),s)$.

Corollary 2.9. *Let $\mathcal{M}$ be a manifold and $X(\mathcal{M})$ be a complete vector field on $\mathcal{M}$ with flow F_t. F_t for $t \in \mathbb{R}$ then forms a group of diffeomorphisms of $\mathcal{M}$ known as **one-parameter group of diffeomorphisms** $\mathrm{Diff}(\mathcal{M})$.*

The corollary also holds in the case of a time-dependent vector field. Diffeomorphism groups are central to continuum mechanics in idealized settings and we will consider them in more detail in the sequel.

Proposition 2.46. *A smooth vector field $X(\mathcal{M})$ with compact support on a manifold $\mathcal{M}$ is complete. For compact $\mathcal{M}$ any smooth vector field is complete.*

Smoothness of a compact vector field $X(\mathcal{M})$ for a non-compact manifold requires $X(\mathcal{M})$ either to have global support or to taper off smoothly so that also integral curves have finite support and a test particle for $X(\mathcal{M})$ would come to rest eventually. For compact $\mathcal{M}$ the support might again not be all of $\mathcal{M}$. A famous example is the "hairy ball theorem" that states that any smooth vector field $X \in S^2$ on the two-sphere has a critical point p where $X(p) = 0$.

Vector Fields as Differential Operators In Def. 2.60, the directional derivative in a vector space was defined as the differential $\mathrm{d}f(\bar{x})$ acting on a vector $X(\bar{x})$. This idea can be generalized to manifolds. However, first we need to introduce a dual space to the tangent bundle.

Definition 2.94. *Let $\mathcal{M}$ be a manifold with tangent space $T_m\mathcal{M}$ at $m \in \mathcal{M}$. The **cotangent space** $T_m^*\mathcal{M}$ **at** m is the dual space of $T_m\mathcal{M}$. The **cotangent bundle** is the disjoint union of all cotangent spaces,*

$$T^*\mathcal{M} = \bigcup_{m \in \mathcal{M}} T_m^*\mathcal{M}.$$

*A **covector field** $\alpha \in \Gamma(T^*\mathcal{M})$ is a section of the cotangent bundle $T^*\mathcal{M}$ and the space of all covector fields is denoted as $\mathfrak{X}^*(\mathcal{M})$. The pairing*

$$\langle\,,\rangle : \mathfrak{X}^*(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to \mathcal{F}(\mathcal{M})$$

between $\mathfrak{X}^*(\mathcal{M})$ and $\mathfrak{X}(\mathcal{M})$ is defined fiber-wise by $\alpha_m(X_m)$ for $\alpha_m \in T_m^*\mathcal{M}$ and $X_m \in T_m\mathcal{M}$.

For $f : \mathcal{M} \to \mathbf{F}$, where $\mathbf{F}$ is a Banach space, Def. 2.94, can be extended to $\mathbf{F}$-valued covector field.

Example 2.45. *Let $\mathcal{M}$ be a manifold and $f : \mathcal{M} \to \mathbb{R}$ be a function defined on $\mathcal{M}$. The differential $\mathrm{d}f$, defined fiber-wise over $\mathcal{M}$, forms a covector field.*

Remark 2.53. Let $\mathcal{M}$ be a manifold and $f \in \mathcal{F}(\mathcal{M})$, and let (U, φ) with $\varphi : U \subset \mathcal{M} \to V \subset \mathbb{R}^n$ be a chart for $\mathcal{M}$. The local representative of f with respect to (U, φ) is the map $f_{\mathrm{lr}} = f \circ \varphi^{-1} : V \to \mathbb{R}$. The local representative $\mathrm{d}f_{\mathrm{lr}}$ of the differential $\mathrm{d}f$ is then

$$\mathrm{d}f_{\mathrm{lr}} = \left(\frac{\partial f_{lr}}{\partial x_1}, \ldots, \frac{\partial f_{lr}}{\partial x_n} \right) \tag{2.84}$$

and it acts on the local representative X_{lr} of a vector field $X \in \mathfrak{X}(\mathcal{M})$. As usual, a differential $\mathrm{d}f$ and its representative $\mathrm{d}f_{\mathrm{lr}}$ are typically identified.

A diffeomorphism between manifolds induces a natural diffeomorphism between their cotangent bundles, and this notion is introduced next.

Definition 2.95. *Let $\mathcal{M}, \mathcal{N}$ be manifolds and $\varphi : \mathcal{M} \to \mathcal{N}$ be a diffeomorphism. The* **cotangent lift** *$T^*\varphi : T^*\mathcal{N} \to T^*\mathcal{M}$ is the map satisfying*

$$\langle (T^*\varphi)\, \alpha, X \rangle = \langle \alpha, (T\varphi)\, X \rangle$$

where $X \in T_m\mathcal{M}$, $\alpha \in T_n^*\mathcal{N}$ with $n = \varphi(m)$, and $\langle\, , \rangle$ is the natural pairing between the tangent and cotangent bundle.

Pointwise, the cotangent lift hence relates the cotangent space $T_{\varphi(m)}^*\mathcal{N}$ of $\mathcal{N}$ at $n = \varphi(m)$ with $T_m^*\mathcal{M}$ of $\mathcal{M}$.

In Def. 2.57 and Def. 2.60 the directional derivative was defined as $\mathrm{d}f(X)$ where $\mathrm{d}f$ is the differential of the function f. Interchanging operands when the differential acts on vectors, $\mathrm{d}f(X) = X \cdot \mathrm{d}f \equiv X[f]$, we can consider X as acting on functions f. This leads to a generalization of the directional derivative to arbitrary manifolds.

Definition 2.96. *Let $X \in \mathfrak{X}(\mathcal{M})$ be a vector field on a manifold $\mathcal{M}$. The* **directional derivative** *$X[f] : \mathcal{F}(\mathcal{M}) \to \mathcal{F}(\mathcal{M})$ of $f \in \mathcal{F}(\mathcal{M})$ along X at $m \in \mathcal{M}$ is*

$$X\,[f]\,(m) = \mathrm{d}f(m)\,(X(m)) = \mathrm{d}f(m) \cdot X(m). \tag{2.85}$$

The directional derivative $X[f] : \mathcal{F}(\mathcal{M}) \to \mathcal{F}(\mathcal{M})$ hence defines a differential operator acting on functions $f \in \mathcal{F}(\mathcal{M})$. Later on, the directional derivative will serve as model for the Lie derivative.

Remark 2.54. Combining our previous remarks about the local representatives for vector fields and differentials for a finite model space $\mathbb{R}^n$, Remark 2.51 and 2.53, the local representative of the directional derivative is

$$X[f] = \sum_{i=1}^n \frac{\partial f}{\partial x^i} X^i \equiv X_{\mathrm{lr}}[f_{\mathrm{lr}}] = \sum_{i=1}^n \frac{\partial f_{\mathrm{lr}}}{\partial x_i} X_{\mathrm{lr}}^i. \tag{2.86}$$

Remark 2.55. A derivation $\mathfrak{D}$ on a ring[103] is an additive ring endomorphism that satisfies the Leibniz identity

$$\mathfrak{D}(fg) = \mathfrak{D}(f)\,g + f\,\mathfrak{D}(g),$$

where f, g are elements in R. An example for a derivation is the usual derivative over the polynomial ring on $\mathbb{R}$. Above we used a vector field to introduce a differential operator. The converse also holds, that is, any derivation on functions on a manifold defines a vector field, and in the finite dimensional case every derivation can be uniquely identified with a differential operator $X[\cdot]$ defined by the vector field $X(\mathcal{M})$. Intuitively, this states that all reasonable differential operators correspond to a vector field.

Remark 2.56. In Remark 2.46, we discussed that for an embedded, finite dimensional manifold a basis $\{\partial/\partial x^1, \ldots, \partial/\partial x^n\}$ for the tangent space $T_m\mathcal{M}$ is induced by a basis $\{e_1, \ldots, e_n\}$ for the model space by $\partial/\partial x^i = (T\varphi^{-1}(m))\,e_i$, where (U, φ) is some chart. The notion can be generalized to any finite dimensional manifold, without reference to an embedding, as follows. Let $f \in \mathcal{F}(U)$ and (U, φ) be a chart for an n-dimensional manifold $\mathcal{M}$, and define n derivations on $\mathcal{F}(U)$ using the local coordinates $x^i : U \to \mathbb{R}$ by

$$\frac{\partial}{\partial x^i}\,(f) = \frac{\partial f}{\partial x^i} = \frac{\partial(f \circ \varphi^{-1})}{\partial x^i} \circ \varphi. \tag{2.87}$$

It can be shown that the n derivations are linearly independent and thus by Remark 2.55 these define n linearly independent vector fields on U that provide *local* bases for $T\mathcal{M}\,|_U$. Moreover, for a vector field X whose local representative has coordinates X^i one has

$$X[f] = \sum_i X^i \frac{\partial f}{\partial x^i} = \left(\sum_i X^i \frac{\partial}{\partial x^i}\right) f, \tag{2.88}$$

cf. Remark 2.54, and hence any vector field is locally represented by

$$X = \sum_i X^i \frac{\partial}{\partial x^i} \in T\mathcal{M}. \tag{2.89}$$

With the basis $\{\partial/\partial x^1, \ldots, \partial/\partial x^n\}$ for the tangent bundle $T\mathcal{M}\,|_U$ in a chart neighborhood $U \subset \mathcal{M}$, a local basis $\{dx^1, \ldots, dx^n\}$ for the cotangent bundle $T^*\mathcal{M}\,|_U$ is defined by the biorthogonality condition

$$dx^i\left(\frac{\partial}{\partial x^j}\right) = \delta_j^i, \tag{2.90}$$

where $\alpha(v)$ denotes the canonical pairing between the primal and dual space. Any covector $\alpha \in T^*\mathcal{M}$ can hence be written as

$$\alpha = \sum_i \alpha_i\,dx^i, \tag{2.91}$$

---

[103]Recall that a ring is an algebraic structure with two binary operations $(+, \cdot)$, usually denoted by addition and multiplication.

where the coefficients α_i are defined as usual by $\alpha(\partial/\partial x^i)$. In particular, the differential $\mathrm{d}f$ of a function $f \in \mathcal{F}(U)$ is hence given by

$$\mathrm{d}f = \sum_i \frac{\partial f}{\partial x^i} dx^i. \tag{2.92}$$

Example 2.46. Consider again the 2-sphere S^2. Let $y_{lm} : S^2 \to \mathbb{R}$ be an arbitrary spherical harmonic, cf. Sec. 2.2.5. Legendre spherical harmonics are usually specified using their expression $y_{lm}(\theta, \phi)$ in spherical coordinates, that is they are specified using their local representative with respect to spherical coordinates, cf. Example 2.35. The local representative of the differential $\mathrm{d}y_{lm} \in T^*S^2$ is hence

$$\mathrm{d}y_{lm} = \left(\frac{\partial y_{lm}}{\partial \theta}, \frac{\partial y_{lm}}{\partial \phi} \right)^T. \tag{2.93}$$

The directional derivative of a Legendre spherical harmonic in azimuthal direction is therefore

$$X\left[y_{lm} \right] = \mathrm{d}y_{lm} \cdot X = \left(\frac{\partial y_{lm}}{\partial \theta}, \frac{\partial y_{lm}}{\partial \phi} \right)^T \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{\partial y_{lm}}{\partial \phi}. \tag{2.94}$$

An important operator extedning the directional derivative, and which will also provide the model for the Lie bracket for vector fields in the following, is the Jacobi-Lie bracket.

Definition 2.97. *Let $\mathcal{M}$ be a manifold and $X, Y \in \mathfrak{X}(\mathcal{M})$. The **Jacobi-Lie bracket** $[X, Y] : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$ is the unique vector field acting as differential operator $[X, Y] : \mathcal{F}(\mathcal{M}) \to \mathcal{F}(\mathcal{M})$ as*

$$[X, Y] = X[Y[\cdot]] - Y[X[\cdot]].$$

By definition the Jacobi-Lie bracket is the commutator for vector fields and it is hence anti-symmetric satisfying $[X, Y] = -[Y, X]$.

Remark 2.57. For a finite dimensional manifold $\mathcal{M}$ and $X, Y \in \mathfrak{X}(\mathcal{M})$, the components $[X, Y]^i$ of the Jacobi-Lie bracket of two vector fields in local coordinates are given by

$$[X, Y]^j = \sum_i X^i \frac{\partial Y^j}{\partial x^i} - Y^i \frac{\partial X^j}{\partial x^i}.$$

More properties of the Jacobi-Lie bracket will be discussed in the context of general Lie brackets in Chapter 2.3.3. Using the Jacobi-Lie bracket we can complete our discussion of submanifolds.

Definition 2.98. *Let $\mathcal{M}$ be a manifold and $E \subset T\mathcal{M}$ be a subbundle of its tangent bundle.*
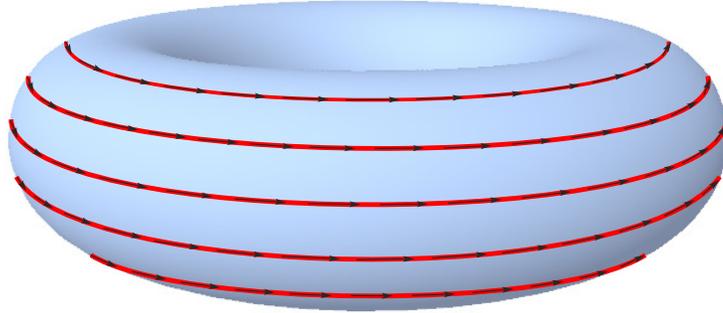
Figure 2.30: Integral manifold (red) for a horizontal vector field (black) on the two-torus $\mathbb{T}^2$.

  i) E is ***involutive*** *if* $X, Y \in E$ *implies* $[X, Y] \in E$.

 ii) E is ***integrable*** *if for any* $m \in \mathcal{M}$ *there is a submanifold* $\mathcal{N} \subset \mathcal{M}$ *containing* m*, the* ***integral manifold*** *of* E *at* m*, whose tangent space is exactly* E *restricted to* $\mathcal{N}$*.*

Note that $\mathcal{N}$ is in general only locally defined.

Example 2.47. Let $\mathbb{T}^2 = S^1 \times S^1$ be the two-torus and $E = \{(0, 1) \in T\mathbb{T}^n\}$. Then E is involutive and integrable and $S^1 \times [0, 2\pi]$ is the integral manifold of E, see Fig. 2.30. The tensor product structure arises since every fiber S^1 is an integral manifold. The example generalizes to the n-torus $\mathbb{T}^n$.

Integrability is a central research direction in the modern theory of differential equations. Integrable systems are those that, in some sense, admit a closed form solution for all times. However, the subject will not be of primary concern to our endeavour.

An important concept in the context of geometric mechanics is the fiber derivative that establishes a correspondence between the tangent and cotangent bundle of a manifold

Definition 2.99. *The* ***fiber derivative*** $\mathbb{F} : TQ \to T^*Q$ *defined by a function* $G : T^*Q \to \mathbb{R}$ *is*

$$\mathbb{F}G(v) \cdot w = \frac{d}{d\epsilon} G(v + \epsilon\, w)$$

for all vectors $v, w \in T_qQ$*, and where the pairing on the left hand side is the natural pairing between* TQ *and* T^*Q*.*

A comparison to Def. 2.61 and the foregoing discussion shows that the fiber derivative can be interpreted as the gradient of $G : TQ \to \mathbb{R}$ in the fiber.

Remark 2.58. In coordinates, the fiber derivative is given by

$$\mathbb{F}G(q^i, \dot{q}^i) = \left(q^i, \frac{\partial G}{\partial \dot{q}^i}\right) \tag{2.95}$$

which shows that $\mathbb{F}$ is fiber-preserving.

2.3.2.3 Tensors

Tensors are linear maps that provide a consistent representation of physical quantities independent of coordinate systems and charts. We already encountered two types of tensors, vectors and covectors, and these will serve as model for contravariant and covariant tensors, sections of vector bundles constructed from the tangent and cotangent bundle. Tensors will first be introduced for linear spaces, which is their natural setting as linear maps, and then be generalized to manifolds, where each fiber of the tangent and cotangent bundle is a linear space.

Tensors on Linear Spaces A covector is an element in $\mathbf{E}^*$, the space of linear maps $L(\mathbf{E}, \mathbb{R})$ from $\mathbf{E}$ to the real numbers $\mathbb{R}$. Conversely, with the pairing between a space and its dual a vector can be regarded as a map $L(\mathbf{E}^*, \mathbb{R})$ from $\mathbf{E}^*$ to $\mathbb{R}$. It is hence natural to consider "higher order" maps that act on multiple elements in $\mathbf{E}$ and $\mathbf{E}^*$.

Definition 2.100. *Let $\mathbf{E}$ be a vector space. A **tensor***

$$t_s^r : \underbrace{\mathbf{E}^* \times \ldots \times \mathbf{E}^*}_{r \text{ times}} \times \underbrace{\mathbf{E} \times \ldots \times \mathbf{E}}_{s \text{ times}} \to \mathbb{R}$$

*on $\mathbf{E}$ of **type** (r, s), **contravariant** of order r and **covariant** of order s, is a map in*

$$T_s^r(\mathbf{E}) = L^{r+s}(\underbrace{\mathbf{E}^*, \ldots, \mathbf{E}^*}_{r \text{ times}}, \underbrace{\mathbf{E}, \ldots, \mathbf{E}}_{s \text{ times}}, \mathbb{R})$$

that is linear in each of its arguments.

The definition generalizes vectors and covectors, as maps acting on their duals, to objects that act on s vectors and r covectors. Tensors where both r and s are nonzero are sometimes denoted as mixed tensors, and a contravariant tensor is an element in T_0^r while a covariant tensor is an element in T_s^0.

Example 2.48. An element of $\mathbf{E}$ is a $(1, 0)$ tensor, that is $T_0^1(E) = E$. Conversely, an element of $\mathbf{E}^*$ is a $(0, 1)$ tensor, that is $T_1^0 = \mathbf{E}^*$.

Remark 2.59. Traditionally, index notation was used almost exclusively for tensors and it is still prevalent in the physics literature. We will embrace modern, index free notation, that is more suited to understand the structure and semantics, whenever possible. However, for practical calculations and computer implementations index notation is needed.

A basis for the space of (r, s) tensors $T_s^r(\mathbf{E})$ is provided by the (r, s)-fold tensor product of bases for $\mathbf{E}$ and $\mathbf{E}^*$.

Definition 2.101. *Let $\mathbf{E}$ be finite dimensional with basis $\{e_1, \ldots, e_n\}$ and dual basis $\{e^1, \ldots, e^n\}$. The **components of a tensor** $t \in T_s^r(\mathbf{E})$ with respect to the bases are*

$$t_{i_1, \cdots, i_s}^{j_1, \cdots, j_r} = t(e^{j_1}, \ldots, e^{j_r}, e_{i_1}, \ldots, e_{i_s}).$$

and the basis representation of the tensor is

$$t = \sum_{i_1 \cdots i_s} \sum_{j_1 \cdots j_r} t_{i_1, \cdots, i_s}^{k_1, \cdots, k_r} \, e_{j_1} \otimes \ldots \otimes e_{j_r} \otimes e^{i_1} \otimes \ldots \otimes e^{i_s}.$$

For $(1, 0)$ and $(0, 1)$ tensors this definition coincides with the usual representation of vectors and covectors. The components of a tensor can hence be regarded as generalized basis coefficients.

Remark 2.60. As pointed out previously, the correct usage of "upstairs" and "downstairs" indices is critical when one works with tensors, and it should be clear now that the reason is to distinguish the contravariant and covariant components of a tensor. Note that the basis vectors $\partial/\partial x^i$ for the tangent space $T\mathcal{M}$ of a manifold $\mathcal{M}$ are considered to have "downstairs" indices so that they naturally pair with the dual basis functions dx^i that have "upstairs" indices, cf. Remark 2.46.

Example 2.49. Let bases for $\mathbf{E}$ and $\mathbf{E}^*$ be given by $\{e_1, \ldots, e_n\}$ and $\{e^1, \ldots e^n\}$. A $t_0^2 \in T_0^2(\mathbf{E})$ tensor then has components $t^{ij} = t(e^1, e^2)$, which allows them to be arranged in matrix form. For example, a $(2, 0)$ tensor on $\mathbb{R}^2$ is given in components by

$$t_0^2 = a^{11} e_1 \otimes e_1 + a^{12} e_1 \otimes e_2 + a^{21} e_1 \otimes e_2 + a^{22} e_2 \otimes e_2, \qquad (2.96)$$

and can be written as

$$t_0^2 = \begin{bmatrix} a^{11} & a^{12} \\ a^{21} & a^{22} \end{bmatrix}. \qquad (2.97)$$

Remark 2.61. There are three types of tensors, elements in T_0^2, T_2^0, and T_1^1, that yield coordinate expressions with two indices which can be arranged in matrix form. However, only elements in T_1^1 are matrices in the usual sense of linear algebra as mappings $t : \mathbf{E} \to \mathbf{E}$. In tensor notation their coordinates are t_i^j, in contrast to the notation t_{ij} used in linear algebra that we employed in Chapter 2.2 and which from a tensorial perspective corresponds to $(0, 2)$ tensors. The elements of T_0^2, T_2^0 are bilinear forms on $\mathbf{E}$ and $\mathbf{E}^*$, respectively. In the following we shall be careful to distinguish elements in the different space T_0^2, T_2^0, T_1^1, even if these can be identified using additional structure such as a metric that is often available in the applications of interest to us.

Definition 2.102. *The **tensor product** of tensors $t_{s_1}^{r_1} \in T_{s_1}^{r_1}$ and $\bar{t}_{s_2}^{r_2} \in T_{s_2}^{r_2}$ is the tensor $\tilde{t}_{s_1+s_2}^{r_1+r_2} \in T_{s_1+s_2}^{r_1+r_2}$ acting as*

$$t_{s_1}^{r_1} \otimes \bar{t}_{s_2}^{r_2}(\alpha_1, \ldots, \alpha_{s_1}, \beta_1, \ldots, \beta_{s_2}, u_1, \ldots, u_{r_1}, v_1, \ldots, v_{r_2})$$
$$= t_{s_1}^{r_1}(\alpha_1, \ldots, \alpha_{s_1}, u_1, \ldots u_{r_1}) \, \bar{t}_{s_2}^{r_2}(\beta_1, \ldots, \beta_{s_2}, v_1, \ldots v_{r_2}).$$

on $\alpha_i, \beta_i \in \mathbf{E}$ and $u_i, v_i \in \mathbf{E}^*$.

In a tensor product, the first r_1 and s_1 arguments are applied to $t_{s_1}^{r_1}$ and the remaining ones to $\bar{t}_{s_2}^{r_2}$, and the result for $t_{s_1}^{r_1}$ and $t_{s_2}^{r_2}$ is multiplied together. The tensor product is associative, bilinear, and continuous, but not commutative.

Remark 2.62. Permutations are important for working with tensors and we will briefly recall some elementary notions. A permutation σ is a bijection $(1, \ldots, k) \to (\sigma(1), \ldots, \sigma(k))$ that rearranges the k elements in its domain. It is often written as

$$\begin{pmatrix} 1 & \cdots & k \\ \sigma(1) & \cdots & \sigma(k) \end{pmatrix}.$$

and the set of all such bijections forms a group denoted as S_k. A permutation σ is a **transposition** when exactly two elements are interchanged; it is *odd* when it consist of an odd number of transpositions, while it is *even* when it can be decomposed into an even number of transpositions, and through the group structure these notions are well defined. The sign $\mathrm{sgn}(\sigma)$ of a permutation is

$$\mathrm{sgn}(\sigma) = \begin{cases} 1 & \sigma \text{ is even} \\ -1 & \sigma \text{ is odd} \end{cases}$$

Definition 2.103. *An $(r,0)$ tensor t is **symmetric** when*

$$t(\alpha_1, \ldots, \alpha_n) = t(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(n)})$$

*for all possible permutations σ of elements $\alpha_1, \ldots, \alpha_n \in \mathbf{E}^*$, and it is **anti-symmetric** or **skew symmetric** when*

$$t(\alpha_1, \ldots, \alpha_n) = \mathrm{sgn}(\sigma) \, t(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(n)}).$$

Analogous definitions holds for $(0,s)$ tensors.

A tensor is hence anti-symmetric when interchanging two of its arguments changes its sign. A tensor that is both covariant and anti-symmetric is a differential forms and we will discuss them in detail in Chapter 2.3.2.4.

Remark 2.63. An $(r,0)$ or $(0,s)$ tensor is symmetric if and only if its components are symmetric; conversely, it is anti-symmetric if and only if its components are anti-symmetric. These notions are to be understood in the sense of transpositions of indices.

Example 2.50. An inner product is a linear, symmetric pairing between elements in a (pre) Hilbert space $\mathcal{H}$, cf. Chapter 2.2.2, and it is hence a $(0, 2)$ tensor on $\mathcal{H}$. In tensor calculus, an inner product is known as metric and commonly denoted as g. In coordinates takes the form

$$g = g_{ij}\, e^i \otimes e^j, \tag{2.98}$$

with the components being defined by $g_{ij} = g(e_i, e_j)$, and the associated inverse metric g^{-1} is

$$g^{-1} = g^{ij}\, e_i \otimes e_j, \tag{2.99}$$

with the components g^{ij} being given by the matrix inverse of the g_{ij}, that is

$$\sum_i g_{ij}\, g^{ki} = \delta_j^k \tag{2.100}$$

where δ_j^k is the Kronecker delta.

The interior product makes precise the idea to apply one elements in $\mathbf{E}$ or $\mathbf{E}^*$ to one of the "slots" of a tensor $t \in T_s^r$ but to leave the remaining ones "open".

Definition 2.104. *The **interior product** $\mathrm{i}_u t : T_s^r \to T_{s-1}^r$ **of a vector** u with an (r, s) tensor t is*

$$\mathrm{i}_u t = t(\alpha_1, \ldots, \alpha_r, u, v_1, \ldots v_{s-1}),$$

*for arbitrary $\alpha_i \in \mathbf{E}^*$ and $v_i \in \mathbf{E}$. The **interior product** $\mathrm{i}^\beta t : T_s^r \to T_s^{r-1}$ **of a covector** β with an (r, s) tensor t is*

$$\mathrm{i}_u t = t(\beta, \alpha_1, \ldots, \alpha_{r-1}, v_1, \ldots v_s).$$

A notion closely related but more general than the interior product is those of contraction which combines an arbitrary number of contravariant and covariant "slots". Since for a mixed (r, s) tensor any pair of indices can be contracted, a family of contraction maps exists.

Definition 2.105. *The (p, q)**-contraction** of the pth contravariant with the qth covariant index is the family of maps $C_q^p : T_s^r \to T_{s-1}^{r-1}$ defined by*

$$C_l^k(t_{i_1,\cdots,i_s}^{k_1,\cdots,k_r}\, e_1 \otimes \ldots \otimes e_r \otimes e^1 \otimes \ldots \otimes e^s)$$

$$= t_{i_1,\cdots,i_{q-1},j,i_{q+1},\cdots,i_s}^{k_1,\cdots,k_{p-1},j,k_{p+1},\cdots,k_r}\, e_1 \otimes \ldots \otimes \hat{e}_{k_p} \otimes \ldots \otimes e_r \otimes e^1 \otimes \ldots \otimes \hat{e}^{i_q} \otimes \ldots \otimes e^s)$$

where summation over j is implied and $\hat{e}_{k_p}, \hat{e}^{i_q}$ represents that these basis vectors are omitted from the tensor product.

Despite our definition, it can be shown that contraction is intrinsic and does not depend on the basis. To define contraction in the infinite dimensional case much care is required and the concept of a contraction class operator, analogous to a trace class operator, has to be employed.

Example 2.51. Let $t \in T_1^1$, cf. Remark 2.61. The **trace** $\mathrm{tr}\,(t)$ of t is the contraction $C_1^1(t)$.

Remark 2.64. The **Einstein summation convention** exploits the use of "upstairs" and "downstairs" indices to imply the summation in a contraction rather than writing it explicitly as we did so far,

$$q_j^k = t_j^i\, p_i^k \equiv \sum_i t_j^i\, p_i^k \tag{2.101}$$

for compatible tensors t_i^j, p_j^k. Unless otherwise noted, in the following we will employ the summation convention.

Remark 2.65. Let $\mathbf{E}$ be finite dimensional with basis $\{e_1, \ldots, e_n\}$ and dual basis $\{e^1, \ldots, e^n\}$. Let $g = g_{ij}\, e^i \otimes e^j \in T_2^0(\mathbf{E})$ be a metric with components g_{ij}, cf. Example 2.50, $X = X^i e_i \in T_0^1(\mathbf{E})$ be a vector, and $\alpha = \alpha_i\, e^i \in T_1^0(\mathbf{E})$ be a covector. Using contraction with g_{ij}, the metric can be employed to associate a covector with X by **raising indices**

$$X_j = g_{ij}\, X^i; \tag{2.102}$$

and a vector can be associated with α using the inverse metric g^{ij} by **lowering indices**

$$\alpha^j = g^{ij} \alpha_i. \tag{2.103}$$

We will have more to say about raising and lowering indices when we discuss the musical isomorphisms in Def. 2.123.

Next, we will study how tensors behave under mappings, which will lead to the notions of push-forward and pullback that, when carried over to the nonlinear case of manifolds, are central for mechanics.

Definition 2.106. *Let $\mathbf{E}, \mathbf{F}$ be Banach spaces and $\mathbf{E}^*, \mathbf{F}^*$ their dual spaces. The **dual** $\varphi^* : \mathbf{F}^* \to \mathbf{E}^*$ of the linear map $\varphi : \mathbf{E} \to \mathbf{F}$, for $u \in \mathbf{E}$ and $\beta \in \mathbf{F}^*$, is*

$$\beta \cdot \varphi(u) = \varphi^*(\beta) \cdot u.$$

The dual is a map in the "inverse direction" of φ, cf. Fig. 2.31. Clearly, when $\mathbf{E}, \mathbf{F}$ are Hilbert spaces, the pairings in the above definitions can be replaced by the inner product using the Riesz representation theorem.

Remark 2.66. Let $\mathbf{E}, \mathbf{F}$ be finite dimensional spaces with bases $\{e_1, \ldots, e_n\}$ and $\{f_1, \ldots, f_n\}$, and let A be the matrix representation of the map $\varphi : \mathbf{E} \to \mathbf{F}$ with respect to these bases. According to Remark 2.61, the components A_i^j of A are given by $\varphi(e_i) = A_i^j f_j$ where j is the row index and i the column index of the matrix. For an arbitrary $v = v^i e_i \in \mathbf{E}$ we hence have

$$\varphi(v) = \varphi(v^i e_i) = v^i \varphi(e_i) = v^i (A_i^j f_j) = (v^i A_i^j) f_j = \bar{v}^j f_j, \tag{2.104}$$
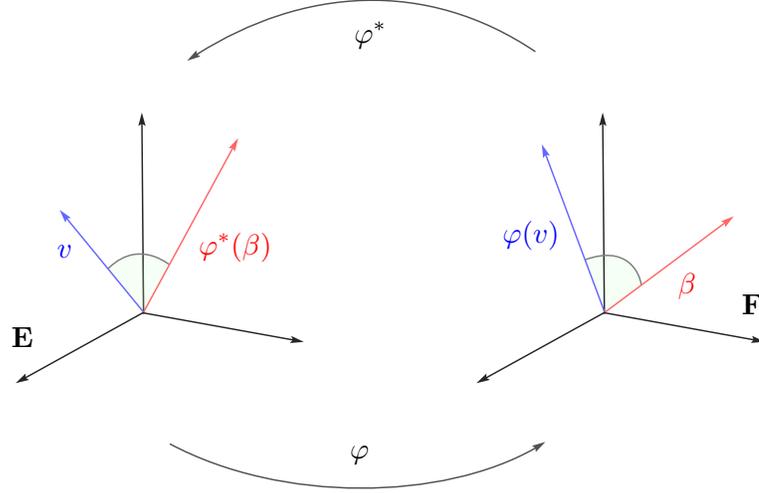
Figure 2.31: The dual $\varphi^* : \mathbf{F}^* \to \mathbf{E}^*$ maps $\beta \in \mathbf{F}^*$ to $\varphi^*(\beta) \in \mathbf{E}^*$ such that the pairing (green) with $v \in \mathbf{E}$ and its image $\varphi(v) \in \mathbf{F}$ is in both spaces identical.

where we used the linearity of φ, and the $\bar{v}^j = v^i A_i^j$, that are obtained by *left* multiplying the vector $(v^1, \ldots, v^n)$ with the matrix A_i^j, are the basis function coefficients of the image of v with respect to $\{f_1, \ldots, f_n\}$. Conversely, for a covector $\beta = \beta_j f^j \in T_1^0(\mathbf{F})$ we have

$$\beta(\varphi(v)) = (\beta_j f^j)(v^i A_i^l f_l) = \beta_j v^i A_i^l f^j(f_l) = v^i A_i^j \beta_j = v^i (A_i^j \beta_j), \qquad (2.105)$$

where we exploited the biorthogonality of the basis functions, $f^j(f_l) = \delta_l^j$. It follows from the last equality that in coordinates $\varphi^* \beta = (A_i^j \beta_j) e^i$ and for the dual φ^* the components $(\beta_1, \ldots, \beta_n)$ of a covector are *right* multiplied by the matrix representation A_i^j of the map φ. For the components of vectors $v = v^i e_i \in \mathbf{E}$ and covectors $\beta = \beta_j f^j \in \mathbf{F}^*$ we hence have the following transformation rules under a map $\varphi : \mathbf{E} \to \mathbf{F}$ with dual $\varphi^* : \mathbf{F}^* \to \mathbf{E}^*$,

$$\varphi(u) \ : \ u^i A_i^j \qquad\qquad \varphi^*(\beta) \ : \ A_i^j \beta_j. \qquad (2.106)$$

A linear transformation $\varphi : \mathbf{E} \to \mathbf{F}$ induces an associated map on tensors by acting on its argument.

Definition 2.107. *Let $\varphi : \mathbf{E} \to \mathbf{F}$ be an isomorphism. The **push-forward** $\varphi_* t : T_s^r(\mathbf{E}) \to T_s^r(\mathbf{F})$ of an (r, s) tensor $t \in T_s^r(\mathbf{E})$ is*

$$(T\varphi)t = \varphi_* t = t(\varphi^* \beta_1, \ldots, \varphi^* \beta_r, \varphi^{-1}(u_1), \ldots, \varphi^{-1}(u_s))$$

for arbitrary $\beta_i \in \mathbf{F}^*$ and $u_i \in \mathbf{F}$.

Analogous to the dual in Def. 2.106, the push-forward of a tensor $t \in T_s^r(\mathbf{E})$ is a tensor $\varphi_* t \in T_s^r(\mathbf{F})$ on $\mathcal{F}$ that agrees in its value with the original tensor t when the arguments are mapped backed from $\mathcal{F}$ to $\mathbf{E}$. The strong requirement for φ to be an isomorphism is necessary since the inverse φ^{-1} has to exist to relate $\mathbf{F}$ and $\mathbf{E}$.

Remark 2.67. It follows from Def. 2.106 that the push-forward of a vector $u \in E$ by $\varphi : \mathbf{E} \to \mathbf{F}$ is $\varphi_* u = \varphi(u)$.

The following proposition summarizes the properties of the push-forward and shows that it is well behaved.

Proposition 2.47. *Let $\varphi : \mathbf{E} \to \mathbf{F}$ and $\psi : \mathbf{F} \to \mathbf{G}$ be isomorphisms. Then*

i) $(\varphi \circ \psi)_* = \varphi_* \circ \psi_*$;

ii) for the identity map $i : \mathbf{E} \to \mathbf{E}$ on $\mathbf{E}$ also the push-forward is the identity;

iii) $\varphi_* : T_s^r(\mathbf{E}) \to T_s^r(\mathbf{F})$ is an isomorphism;

iv) $(\varphi_*)^{-1} = (\varphi^{-1})_*$;

v) for $t_1 \in T_{s_1}^{r_1}$ and $t_2 \in T_{s_2}^{r_2}$, $\varphi_*(t_1 \otimes t_2) = \varphi_*(t_1) \otimes \varphi_*(t_2)$.

The "inverse push-forward" $(\varphi^{-1})_*$ plays a central role for covariant tensors and is of interest in its own right.

Definition 2.108. *Let φ_* be the push-forward by the map $\varphi : \mathbf{E} \to \mathbf{F}$. The **pullback** is the map $\varphi^* \equiv (\varphi^{-1})_* = T_s^r(\mathbf{F}) \to T_s^r(\mathbf{E})$ given by*

$$\varphi^* t = t \left((\varphi^{-1})^* \alpha_1, \ldots, (\varphi^{-1})^* \alpha_r, \varphi(v_1), \ldots, \varphi(v_s) \right)$$

for arbitrary $\alpha_i \in \mathbf{E}^*$ and $v_i \in \mathbf{E}$.

Remark 2.68. Let $\mathbf{E}$ and $\mathbf{F}$ be finite dimensional Banach spaces with bases $\{e_1, \ldots, e_n\}$ and $\{f_1, \ldots, f_n\}$, respectively, and let $\varphi : \mathbf{E} \to \mathbf{F}$ be an isomorphism whose basis representation is the matrix A with elements A_i^j. The matrix representation of φ^{-1} is then the matrix B with elements B_k^i such that $A_i^j B_k^i = \delta_k^j$, that is B is the matrix inverse of A. With Eq. 2.106, the push-forward of a tensor $t \in T_s^r(\mathbf{E})$ with components $t_{i_1, \cdots, i_s}^{j_1, \cdots, j_r}$ with respect to $\{e_1, \ldots, e_n\}$ is thus

$$(\varphi_* t)_{k_1, \cdots k_s}^{l_1, \cdots l_r} = A_{j_1}^{l_1} \ldots A_{j_r}^{l_r} \, t_{i_1, \cdots, i_s}^{j_1, \cdots, j_r} B_{k_1}^{i_1} \ldots B_{k_s}^{i_s} \tag{2.107}$$

where the $(\varphi_* t)_{k_1, \cdots k_s}^{l_1, \cdots l_r}$ are the components with respect to $\{f_1, \ldots, f_n\}$. By linearity of t, each argument is transformed individually and hence r copies of A and s copies of B appear in the above formula. Correspondingly, the coordinate expression of the pullback for a tensor $t \in T_s^r(\mathbf{F})$ with components $t_{k_1, \cdots k_s}^{l_1, \cdots l_r}$ with respect to $\{f_1, \ldots, f_n\}$ is

$$(\varphi^* t)_{i_1, \cdots, i_s}^{j_1, \cdots, j_r} = B_{l_1}^{j_1} \ldots B_{l_r}^{j_r} \, t_{k_1, \cdots k_s}^{l_1, \cdots l_r} A_{i_1}^{k_1} \ldots A_{i_s}^{k_s}. \tag{2.108}$$

Remark 2.69. For $\mathbf{F} = \mathbf{E}$, so that $\varphi : \mathbf{E} \to \mathbf{E}$, the push-forward determines how the coordinate expressions for tensors $t \in T_s^r(\mathbf{E})$ transform under a change of coordinates.

For the push-forward and the pullback of arbitrary tensors, the map φ has to be an isomorphism since we have to employ the inverse to relate the elements of $\mathbf{F}$ to those in $\mathbf{E}$. This is a rather severe restriction on the mapping. However, from Def. 2.107 and Def. 2.108 it follows that we can dispense with this requirement for covariant tensors.

Definition 2.109. *Let $\varphi : \mathbf{E} \to \mathbf{F}$ be a linear map. The **pullback** $\varphi^* : T_s^0(\mathbf{F}) \to T_s^0(\mathbf{E})$ of a covariant tensor $\alpha \in T_s^0(\mathbf{F})$ is*

$$(\varphi^* \alpha)(v_1, \ldots, v_s) = \alpha \left(\varphi(v_1), \ldots, \varphi(v_s) \right)$$

where the v_i are arbitrary elements in $\mathbf{E}$.

In the above definition the inverse φ^{-1} is no longer needed, and the requirement for φ to be an isomorphism has been dropped. Evidently, the definition for the pullback also motivated our notation for the dual of a linear map in Def. 2.106, which is the pullback of an element in $T_1^0(\mathbf{F}) \cong \mathbf{F}^*$. The properties of the pullback for covariant tensors are summarized in the following proposition.

Proposition 2.48. *Let $\varphi : \mathbf{E} \to \mathbf{F}$ and $\psi : \mathbf{F} \to \mathbf{G}$ and consider only covariant tensors in $T_s^0(\mathbf{E})$, $T_s^0(\mathbf{F})$, and $T_s^0(\mathbf{G})$. Then*

i) $(\varphi \circ \psi)^* = \varphi^* \circ \psi^*$;

ii) for the identity map $i : \mathbf{E} \to \mathbf{E}$ on $\mathbf{E}$ also the pullback is the identity;

iii) if φ is an isomorphism then also φ^* is an isomorphism and $\varphi^* = (\varphi^{-1})_*$;

iv) if $t_1 \in T_s^0(\mathbf{F})$ and $t_2 \in T_s^0(\mathbf{F})$ then $\varphi^*(t_1 \otimes t_2) = \varphi^* t_1 \otimes \varphi^* t_2$.

Example 2.52.[104] Let $t \in T_1^1(\mathbb{R}^2)$ be given by $t = e_1 \otimes e^2 - 2e_2 \otimes e^2$ and the matrix representation A of the linear map $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ with respect to the canonical basis $\{e_1, e_2\}$ be

$$A = \left[\begin{array}{cc} 2 & 1 \\ 1 & 1 \end{array} \right]. \tag{2.109}$$

The inverse of A exists, which implies that A is an isomorphism, and it has the representation

$$B = A^{-1} = \left[\begin{array}{cc} 1 & -1 \\ -1 & 2 \end{array} \right]. \tag{2.110}$$

---

[104]The example is based on (Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*, Example 6.1.11.C).

Using the linearity of the push-forward and that it commutes with tensor products, Proposition 2.48, iv), we have for the push-forward,

$$\varphi_* t = \varphi_*(e_1) \otimes \varphi_*(e^2) - 2\varphi_*(e_2) \otimes \varphi_*(e^2). \tag{2.111}$$

As discussed in Remark 2.66 and Remark 2.68, and using Remark 2.67, the push-forward of the basis functions is given by

$$\varphi_*(e_1) = (1,0)\,A = 2e_1 + e_2 \qquad \varphi_*(e^1) = B\,(1,0)^T = e^1 - e^2 \tag{2.112}$$

$$\varphi_*(e_2) = (0,1)\,A = e_1 + e_2 \qquad \varphi_*(e^2) = B\,(0,1)^T = -e^1 + 2e^2 \tag{2.113}$$

and hence

$$\varphi_*(t) = (2e_1 + e_2) \otimes (-e^1 + 2e^2) - 2(e_1 + e_2) \otimes (-e^1 + 2e^2) \tag{2.114}$$

$$= -2e_1 \otimes e^1 + 4e_1 \otimes e^2 - e_2 \otimes e^1 + 2e_2 \otimes e^2 \tag{2.115}$$

$$+ 2e_1 \otimes e^1 - 4e_1 \otimes e^2 + 2e_2 \otimes e^1 - 4e_2 \otimes e^2 \tag{2.116}$$

$$= e_2 \otimes e^1 - 2e_2 \otimes e^2. \tag{2.117}$$

Alternatively, we can compute the push-forward $\varphi_* t$ directly using Eq. 2.107. In coordinates, with respect to the basis $e_i \otimes e^j$, whose basis functions are $\{e_1 \otimes e^1, e_1 \otimes e^2, e_2 \otimes e^1, e_2 \otimes e^2\}$, t is given by $t_i^j = (0,1,0,2)$. In matrix form this reads

$$t_i^j = \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix},$$

cf. Example 2.49. The coordinate expression for the push-forward of a $(1,1)$ tensor is by Eq. 2.107

$$(\varphi_* t)_l^k = A_j^k \, t_i^j \, B_l^i,$$

which again can be written in matrix form,

$$\varphi_* t = A\,t\,B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -2 \end{bmatrix}.$$

Reading off the components then gives $\varphi_* t = e_2 \otimes e^1 - 2e_2 \otimes e^2$, recovering our previous result.

Tensor Bundles After studying tensors on linear spaces we are ready to introduce them for manifolds. We will first define them for local vector bundles and then generalize to manifolds and the tensor bundle. For the next definition it will be useful to have the commuting diagram in Eq. 2.81 in mind and the definition of a local vector bundle map in Def. 2.76.

Definition 2.110. *Let $\varphi : U \times \mathbf{E} \to V \times \mathbf{F}$ be a local vector bundle map that is an isomorphism over each fiber so that $\varphi_{\bar{u}} \equiv \varphi\,|_{\{\bar{u}\} \times \mathbf{E}} : \mathbf{E} \to \mathbf{F}$ is an isomorphism. The push-forward $\varphi_* : U \times T_s^r(\mathbf{E}) \to V \times T_s^r(\mathbf{F})$ of an (r,s) tensor $t \in T_s^r(\mathbf{E})$ is then $\varphi_*(u,t) = (\varphi(u),(\varphi_u)_* t)$.*

It follows from linearity that if $U \times \mathbf{E}$ is a local vector bundle then so it $U \times T_s^r(\mathbf{E})$. It can be shown that when φ_u is a local vector bundle map that is an isomorphism over each fiber then φ_* is a vector bundle map and $(\varphi_*)_u$ is an isomorphism so that the following diagram commutes:

$$
\begin{array}{ccc}
U \times T_s^r(\mathbf{E}) & \xrightarrow{\varphi_*} & V \times T_s^r(\mathbf{F}) \\
\pi_{\mathbf{E}} \downarrow & & \downarrow \pi_{\mathbf{F}} \\
U & \xrightarrow{\varphi_*} & V
\end{array}
$$

Hence the push-forward φ_* as a map between two local vector bundles is smooth and it can be used to define transition maps. Using charts that have locally the structure $U \times T_s^r(\mathbf{E})$ and such transition maps, we can equip the space of tensors on a manifold, defined fiber by fiber, with a manifold structure.

Definition 2.111. *Let $\pi : E \to B$ be a vector bundle and $E_b = \pi^{-1}(b)$ be the fiber over $b \in B$. The (r, s) **tensor bundle** $T_s^r(E)$ of E is the vector bundle*

$$
T_s^r(E) = \bigcup_{b \in B} T_s^r(E_b)
$$

*with fiber projection $\pi_s^r : T_s^r(E) \to B$ given by $\pi_s^r(t) = b$ for $t \in T_s^r(E_b)$. For another vector bundle $\pi' : E' \to B'$ and a vector bundle map $\varphi : E \to E'$, whose restriction to the fiber over $b \in B$ is $\varphi_b = \varphi|_{E_b}$ and that is an isomorphism over each fiber, cf. Eq. 2.81, the **push-forward** $\varphi_* : T_s^r(E) \to T_s^r(E')$ of the tensor bundle $T_s^r(E)$ by φ is defined fiber-wise by $\varphi_*|_{T_s^r(E_b)} = (\varphi_b)_*$*

Note that the vector bundle $T_s^r(E)$ is *not* a bundle over E but its base is again B and the original bundle $\pi : E \to B$ is the special case $T_0^1(E)$. In the above definition the space $T_s^r(E_b)$ of (r, s) tensors on each fiber E_b is well defined since E_b has a natural vector space structure and hence Def. 2.100 applies. The push-forward of a tensor bundle is defined fiber-wise so that Def. 2.107 applies where the push-forward is defined for linear spaces. It can be shown that a vector bundle map $\varphi : E \to E'$ that is an isomorphism on each fiber induces a push-forward $\varphi_* : T_s^r(E) \to T_s^r(E')$ that is also a vector bundle map on the bundles $T_s^r(E)$ and again an isomorphism on fibers $T_s^r(E_b)$. Vector bundle maps that are fiber isomorphisms satisfy the usual properties, analogous to Proposition 2.47.

Remark 2.70. The approach of defining a concept on a manifold fiber-wise using a linear analogue will be seem more often in the following.

Of particular relevance for applications are tensor bundles defined over the tangent bundle $T\mathcal{M}$.

Definition 2.112. *Let $\mathcal{M}$ be a manifold and $\pi : T\mathcal{M} \to \mathcal{M}$ be its tangent bundle. The **tangent tensor bundle** $T_s^r(\mathcal{M}) \equiv T_s^r(T\mathcal{M})$ is the vector bundle*

*of (r,s) tensors **contravariant** of order r and **covariant** of order s. One identifies the tangent bundle with $T\mathcal{M} = T_0^1(\mathcal{M})$ and the cotangent bundle with $T^*\mathcal{M} = T_1^0(\mathcal{M})$.*

Tensor fields are the generalization of functions and (co)vector fields to arbitrary sections of $T_s^r(\mathcal{M})$.

Definition 2.113. *An (r,s) **tensor field** on a manifold is a section of $T_s^r(\mathcal{M})$. The space of all section is denoted by $\mathcal{T}_s^r(\mathcal{M})$ and it has a natural, infinite dimensional vector space structure. We identify $T_0^0(\mathcal{M}) = \mathcal{F}(\mathcal{M})$. Evaluation, scalar multiplication, tensor product, and other operations defined for tensors are fiber-wise defined for tensor fields.*

It is important to distinguish the infinite dimensional space $\mathcal{T}_s^r(\mathcal{M})$, whose elements are sections of $T_s^r(\mathcal{M})$ and which has a vector space structure, from the vector bundle $T_s^r(\mathcal{M})$.

Example 2.53. Let $f \in \mathcal{F}(\mathcal{M})$ and $t \in \mathcal{T}_s^r(\mathcal{M})$ a tensor field. Scalar multiplication of a tensor field is defined fiber-wise by $ft : \mathcal{M} \to \mathcal{T}_s^r(\mathcal{M}) : m \to f(m)\,t(m)$.

Remark 2.71. For manifolds $\mathcal{M}, \mathcal{N}$, a map $\varphi : \mathcal{M} \to \mathcal{N}$ induces a natural hierarchy of mappings

$$\varphi : \mathcal{M} \to \mathcal{N} \tag{2.118a}$$

$$T\varphi : T\mathcal{M} \to T\mathcal{N} \tag{2.118b}$$

$$(T\varphi)_* : T_s^r(\mathcal{M}) \to T_s^r(\mathcal{N}). \tag{2.118c}$$

If φ is a diffeomorphism then $T\varphi$ and $(T\varphi)_*$ are isomorphisms over each fiber.

As already discussed, for $T\mathcal{M}$ a natural chart is induced by the tangent map $T\varphi$ where (U,φ) is an admissible chart for $\mathcal{M}$, and in the finite dimensional case bases for $T_m\mathcal{M}$ and $T_m^*\mathcal{M}$ are given by $\{\partial/\partial x^1, \ldots, \partial/\partial x^n\}$ and $\{dx^1, \ldots, dx^n\}$, see Remark 2.46 and Remark 2.56. Using the push-forward $(T\varphi)_*$ natural charts for $T_s^r(\mathcal{M})$ are also induced by a chart (U,φ) for $\mathcal{M}$.

In Def. 2.101 we introduced the components of a tensor defined over a linear space. We will now generalize this to tensor fields defined over manifolds.

Definition 2.114. *Let $\mathcal{M}$ be a finite dimensional manifold. The **components of a tensor field** $t \in \mathcal{T}_s^r(\mathcal{M})$ with respect to a chart (U,φ) for $\mathcal{M}$ are smooth functions*

$$t_{i_1, \cdot, i_s}^{j_1, \cdots, j^r} = t\left(dx^{i_1}, \ldots, dx^{i_r}, \frac{\partial}{\partial x^{j_1}}, \ldots, \frac{\partial}{\partial x^{j_s}}\right) \in \mathcal{F}(U)$$

where $\{\partial/\partial x^1, \ldots, \partial/\partial x^n\}$ and $\{dx^1, \ldots, dx^n\}$ are the induced bases for $T\mathcal{M}$ and $T^*\mathcal{M}$, respectively. A tensor field on a finite dimensional manifold is therefore locally represented by

$$t\mid_U = t_{i_1, \cdots, i_s}^{j_1, \cdots, j_r} \frac{\partial}{\partial x^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_r}} \otimes dx_{i_1} \otimes \ldots dx_{i_s}.$$
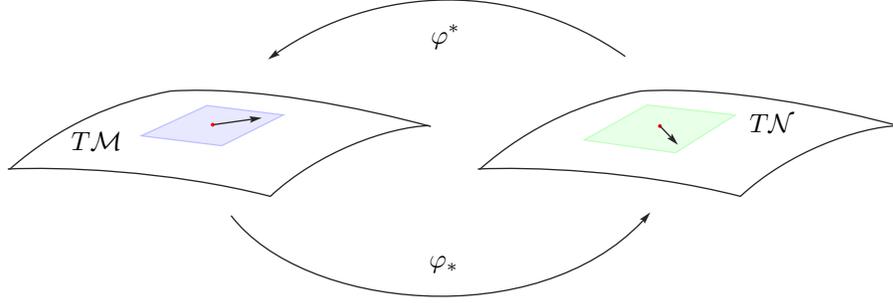
Figure 2.32: Push-forward φ_* and pullback φ^* associated with a map $\varphi : \mathcal{M} \to \mathcal{N}$.

As mentioned before in Remark 2.60, the local basis functions $\partial/\partial x^i$ have, by convention, "downstairs" indices.

Remark 2.72. It is of interest to understand how the components of a tensor with respect to a chart (U, φ) change when a different chart (V, ψ) is employed. The mapping between the charts is given by the transition map $(\psi \circ \varphi^{-1})$: $\mathbb{R}^n \to \mathbb{R}^n$ and the tangent $T(\psi \circ \varphi^{-1})$ provides the required map relating vectors on $\varphi(U) \subset \mathbb{R}^n$ to vectors on $\varphi(V) \subset \mathbb{R}^n$. By Remark 2.45, in coordinates the tangent of the transition map is given by the Jacobian J_i^j associated with the coordinate expression of $(\psi \circ \varphi^{-1})$ and it relates the basis vectors $(\partial/\partial z^1, \ldots, \partial/\partial z^n)$ of V to those of U by $\partial/\partial z^j = J_j^i \partial/\partial x^i$. Using that the tangent $T(\psi \circ \varphi^{-1}) : \varphi(U) \to \psi(V)$ is a linear map between linear spaces, the components $\bar{t}_{k_1, \cdots, k_s}^{l_1, \cdots, l_r}$ of t with respect to (V, ψ) can be obtained using Remark 2.107,

$$\bar{t}_{k_1, \cdots, k_s}^{l_1, \cdots, l_r} = \tilde{J}_{k_1}^{i_1} \ldots \tilde{J}_{k_s}^{i_s} \, t_{i_1, \cdots, i_s}^{j_1, \cdots, j_r} J_{l_1}^{j_1} \ldots J_{l_r}^{j_r}, \qquad (2.119)$$

where $\tilde{J}_i^j$ denotes the inverse of the Jacobian satisfying $J_i^j \tilde{J}_j^k = \delta_i^k$. Historically, Eq. 2.119 was used to *define* a tensor and it is sometimes referred to as **tensorality condition**: A set of $n = r + s$ functions in a local neighborhood $U \subset \mathcal{M}$ of a manifold $\mathcal{M}$ defines an (r, s) tensor if and only if they behave under a change of coordinates according to Eq. 2.119.

We will now generalize push-forward and pullback from the linear setting considered before to tensor fields, see Fig. 2.32.

Definition 2.115. *Let $\varphi : \mathcal{M} \to \mathcal{N}$ be a diffeomorphism between manifolds. Then the **push-forward** $\varphi_* : \mathcal{T}_s^r(\mathcal{M}) \to \mathcal{T}_s^r(\mathcal{N})$ of a tensor field $t \in \mathcal{T}_s^r(\mathcal{M})$ is*

$$\varphi_* t = (T\varphi)_* \circ t \circ \varphi^{-1},$$

*and the **pullback** $\varphi^* : \mathcal{T}_s^r(\mathcal{N}) \to \mathcal{T}_s^r(\mathcal{M})$ of a tensor $t \in \mathcal{T}_s^r(\mathcal{N})$ is*

$$\varphi^* t = (\varphi^{-1})_* t.$$

The inverse map φ^{-1} that determines where the push-forward is evaluated is necessary for pairings with vectors and covectors, that define the meaning of the linear map t_s^r, to be invariant under the mapping. The properties of the push-forward for tensor fields are summarized in the following proposition, see also Proposition 2.47.

Proposition 2.49. *Let $\varphi : \mathcal{M} \to \mathcal{N}$ and $\psi : \mathcal{N} \to \mathcal{K}$ be diffeomorphisms between manifolds, and $t \in \mathcal{T}_s^r(\mathcal{M})$. Then*

i) $\varphi_* t \in \mathcal{T}_s^r(\mathcal{N})$;

ii) $\varphi_* : \mathcal{T}_s^r(\mathcal{M}) \to \mathcal{T}_s^r(\mathcal{N})$ *is a (linear) isomorphism;*

iii) $(\psi \circ \varphi)_* = \psi_* \circ \varphi_*$;

iv) $t_1 \in \mathcal{T}_{s_1}^{r_1}(\mathcal{M}), t_2 \in \mathcal{T}_{s_2}^{r_2}(\mathcal{M})$, *then* $\varphi_*(t_1 \otimes t_2) = \varphi_* t_1 \otimes \varphi_* t_2$.

Two special cases of particular relevance are the push-forward and pullback of functions and vector fields.

Definition 2.116. *Let $\mathcal{M}, \mathcal{N}$ be manifolds and $\varphi : \mathcal{M} \to \mathcal{N}$. The **pullback** $\varphi^* : \mathcal{F}(\mathcal{N}) \to \mathcal{F}(\mathcal{M})$ **of a function** $f \in \mathcal{F}(\mathcal{M})$ by φ is*

$$\varphi^* f = f \circ \varphi.$$

*When the inverse φ^{-1} exists, the **push-forward** $\varphi_* : \mathcal{F}(\mathcal{M}) \to \mathcal{F}(\mathcal{N})$ **of a function** $g \in \mathcal{F}(\mathcal{M})$ by φ is*

$$\varphi_* g = (\varphi^{-1})^* g = g \circ \varphi^{-1}$$

Remark 2.73. The pullback of functions provides an algebra homomorphism between $\mathcal{F}(\mathcal{N})$ and $\mathcal{F}(\mathcal{M})$.

The pullback is well defined when φ exists, the map does not even have to be smooth. However, the push-forward exists only when the inverse φ^{-1} exists.

Definition 2.117. *Let $\mathcal{M}, \mathcal{N}$ be manifolds and $\varphi : \mathcal{M} \to \mathcal{N}$. The **push-forward** $\varphi_* : \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{N})$ of a vector field $X \in \mathfrak{X}(\mathcal{M})$ by φ is*

$$\varphi_* X = T\varphi \circ X \circ \varphi^{-1},$$

*and the **pullback** $\varphi^* : \mathfrak{X}(\mathcal{N}) \to \mathfrak{X}(\mathcal{N})$ of a vector field $Y \in \mathfrak{X}(\mathcal{N})$ by φ is*

$$\varphi^* Y = (\varphi^{-1})_* Y = (T\varphi)^{-1} \circ Y \circ \varphi.$$

Next, we introduce an equivalence relationship for vector fields under mappings that is of particular importance when flows are considered.

Definition 2.118. *Let $\varphi : \mathcal{M} \to \mathcal{N}$ be a smooth mapping between manifolds. The vector fields $X \in \mathfrak{X}(\mathcal{M})$ and $Y \in \mathfrak{X}(\mathcal{N})$ are φ-**related** $X \sim_\varphi Y$ when $(T\varphi)X = Y \circ \varphi$.*

The following proposition connects the above definition to the flow of the vector fields, and when the domain is appropriately restricted the results carry over to local flows.

Proposition 2.50. *Let $\mathcal{M}$ be a manifold and $X \in \mathcal{M}$ be a (not necessarily time invariant) vector field on $\mathcal{M}$ with flow F_t. Then $F_t^* X = X$.*

The above proposition can be generalized to vector fields on different manifolds as follows.

Proposition 2.51. *Let $\varphi : \mathcal{M} \to \mathcal{N}$ be a smooth mapping between manifolds, and let F_t^X and F_t^Y be the flows generated by the vector fields $X \in \mathfrak{X}(\mathcal{M})$ and $Y \in \mathfrak{X}(\mathcal{N})$ on $\mathcal{M}$ and $\mathcal{N}$, respectively. Then $\varphi \circ F_t^X = F_t^Y \circ \varphi$ and φ commutes with the flows if and only if $X \sim_\varphi Y$. If φ is a diffeomorphism then $Y = \varphi_* X$ if and only if F_t^Y is the **push-forward** $F_t^Y = \varphi \circ F_t^X \circ \varphi^{-1}$ **of the flow** F_t^X.*

For a diffeomorphism we hence have that the push-forward of the flow is the flow of the push-forward of the vector field.

Remark 2.74. Let $\mathcal{M}, \mathcal{N}$ be finite dimensional manifolds and $\varphi : \mathcal{M} \to \mathcal{N}$ a diffeomorphism, and let $t \in \mathcal{T}_s^r(\mathcal{M})$ and $\bar{t} \in \mathcal{T}_s^r(\mathcal{N})$. The coordinate expression for the push-forward and pullback follow from those in the linear case in Remark 2.68, applied fiber-wise, with the basis vectors for the tangent spaces being related by the Jacobian associated with φ, cf. Remark 2.72. Hence, when $\partial y^j / \partial x^i$ and $\partial x^k / \partial y^l$ denote the elements of the Jacobian J and inverse Jacobian $\tilde{J}$, respectively, the push-forward of $t \in \mathcal{T}_s^r(\mathcal{M})$ at $n \in \mathcal{N}$ with $m_{\varphi^{-1}} = \varphi^{-1}(n)$ is

$$(\varphi_* t)_{k_1, \cdots, k_s}^{l_1 \cdots l_r}(n)$$
$$= \frac{\partial y^{l_1}}{\partial x^{j_1}}(m_{\varphi^{-1}}) \ldots \frac{\partial y^{l_r}}{\partial x^{j_r}}(m_{\varphi^{-1}}) \, t_{i_1, \cdots, i_s}^{j_1, \cdots, j^r}(m_{\varphi^{-1}}) \frac{\partial x^{i_1}}{\partial y^{l_1}}(n) \ldots \frac{\partial x^{i_s}}{\partial y^{l_s}}(n)$$
$$\tag{2.120}$$

and the pullback of $\bar{t} \in \mathcal{T}_s^r(\mathcal{N})$ at $m \in \mathcal{M}$ with $n_\varphi = \varphi(m)$ is

$$(\varphi^* \bar{t})_{k_1, \cdots, k_s}^{l_1 \cdots l_r}(m)$$
$$= \frac{\partial x^{l_1}}{\partial y^{j_1}}(n_\varphi) \ldots \frac{\partial x^{l_s}}{\partial y^{j_s}}(n_\varphi) \, t_{i_1, \cdots, i_s}^{j_1, \cdots, j^r}(n_\varphi) \frac{\partial y^{i_1}}{\partial x^{k_1}}(m) \ldots \frac{\partial y^{i_r}}{\partial x^{k_r}}(m). \tag{2.121}$$

Analogous to the linear case, the restrictive requirement of φ being a diffeomorphism, an isomorphism in the category of smooth manifolds, can be relaxed when one restricts the pullback to covariant tensors.

Definition 2.119. *Let $\varphi : \mathcal{M} \to \mathcal{N}$ and $t \in T_s^0(\mathcal{N})$ be a covariant tensor on $\mathcal{N}$. The **pullback** $\varphi^* : T_s^0(\mathcal{N}) \to T_s^0(\mathcal{M})$ of t at $m \in \mathcal{M}$ for $v_i \in \mathfrak{X}(\mathcal{M})$ is*

$$(\varphi^* t)(m) = t(\varphi(m))(T_m \varphi(v_1), \ldots, T_m \varphi(v_s)).$$
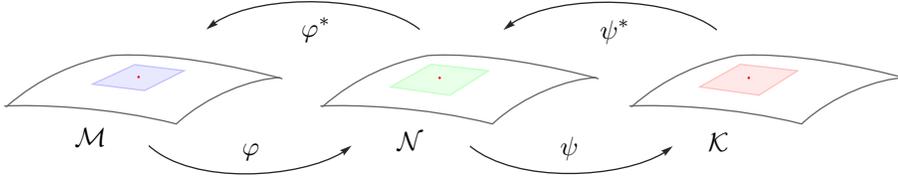
Figure 2.33: The pullback $(\psi \circ \varphi)$ of the composite map $\psi \circ \varphi : \mathcal{M} \to \mathcal{N} \to \mathcal{K}$ is applied in the inverse order so that $(\psi \circ \varphi) = \varphi^* \circ \psi^*$.

The properties of the pullback for covariant tensors are again summarized in a proposition, see also Fig. 2.33.

Proposition 2.52. *Let $\varphi : \mathcal{M} \to \mathcal{N}$ and $\psi : \mathcal{N} \to \mathcal{K}$, and let $t \in \mathcal{T}_s^0(\mathcal{N})$. Then*

i) $\varphi^* t \in \mathcal{T}_s^0(\mathcal{M})$;

ii) $\varphi^* : T_s^0(\mathcal{N}) \to T_s^0(\mathcal{M})$ is a linear map;

iii) $(\psi \circ \varphi)^* = \varphi^* \circ \psi^*$;

iv) if φ is a diffeomorphism then φ^* is an isomorphism with $(\varphi^*)^{-1} = (\varphi^{-1})^*$;

v) $t_1 \in T_{s_1}^0(\mathcal{N}), t_2 \in T_{s_2}^0(\mathcal{N})$, then $\varphi^*(t_1 \otimes t_2) = \varphi^* t_1 \otimes \varphi^* t_2$.

The linearity of the pullback implies that for $\alpha, \beta \in \mathcal{T}_s^r(\mathcal{N})$ and $f, g \in \mathcal{F}(\mathcal{M})$ one has $\varphi^*(f\alpha + g\beta) = f(\varphi^*\alpha) + g(\varphi^*\beta)$, and the natural, infinite dimensional vector space structure of $\mathcal{T}_s^0$ is preserved under the mapping.

Remark 2.75. For finite dimensional manifolds, the coordinate expression for the pullback of a covariant tensor $t \in T_s^0(\mathcal{N})$ by a map $\varphi : \mathcal{M} \to \mathcal{N}$ that is locally given by $y^j = \varphi^j(x^1, \dots x^n)$ follows immediately from our previous remarks,

$$(\varphi^* t)_{j_1, \cdots, j_s}(m) = t_{i_1, \cdots, i_s}(\varphi(m)) \frac{\partial y_{i_1}}{\partial x^{j_1}}(m) \cdots \frac{\partial y_{i_s}}{\partial x^{j_s}}(m), \qquad (2.122)$$

where the $\partial y_{i_l}/\partial x_{i_l}$ are the elements of the Jacobian of φ. Eq. 2.122 is also a special case of Eq. 2.121.

Example 2.54. Let $\mathcal{M}$ be a finite dimensional manifold with model space $\mathbf{E}$, and let a basis for $\mathbf{E}$ be given by $\{e_1, \dots, e_n\}$ with dual basis $\{e^1, \dots, e^n\}$. As discussed in Remark 2.56, in the finite dimensional case a basis for $T_m^* \mathcal{M}$ is given by $\{dx^1, \dots, dx^n\}$ where the dual basis functions dx^i are defined by the biorthogonality condition $dx^i(\partial/\partial x^j) = \delta_j^i$ and the primal basis functions are $\partial/\partial x^i = (T\varphi^{-1}(m)) e_i$. By linearity of the pullback in each fiber, that follows from the fiber-wise linearity of the push-forward, one hence has $dx^i = \varphi^*(e^i)$, so that the dual basis functions dx^i for $T\mathcal{M}$ are given by the pullback of the dual basis functions e^i from the chart.

Example 2.55.[105] Let $\varphi : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$\varphi(x, y, z) = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 2x + z \\ x\,y\,z \end{pmatrix}$$

and $t \in \mathcal{T}_2^0(\mathbb{R}^2)$ be

$$t = (u + 2v)\,du \otimes du + u^2 du \otimes dv.$$

For the pullback we need the Jacobian

$$J_\varphi = \begin{pmatrix} \partial u/\partial x & \partial u/\partial y & \partial u/\partial z \\ \partial v/\partial x & \partial v/\partial y & \partial v/\partial z \end{pmatrix} = \begin{pmatrix} 2 & 0 & 1 \\ yz & xz & xz \end{pmatrix}$$

By Eq. 2.122 we obtain for the basis functions

$$\varphi^*(du) = (1, 0)\,J_\varphi = 2dx + dz$$

$$\varphi^*(dv) = (0, 1)\,J_\varphi = yz\,dx + xz\,dy + xy\,dz$$

which, together with Theorem 2.52 iv), gives

$$\varphi^*t = ((2x + z) + 2(x\,y\,z))\,(2dx + dz) \otimes (2dx + dz)$$
$$+ (2x + z)^2 (2dx + dz) \otimes (yz\,dx + xz\,dy + xy\,dz).$$

A partition of unity is a technical device to smoothly piece together local parts of a tensor field. In the following we will employ it to obtain a globally defined local representative of a tensor field. Later on it will play an important role when we define integration on manifolds.

Definition 2.120. *Let $\mathcal{M}$ be a manifold and $\mathcal{A} = \{(U_j, \varphi_j)\}$ be an atlas for $\mathcal{M}$ A **partition of unity** on $\mathcal{M}$ is a collection $\{(V_i, g_i)\}$ such that*

i) the V_i form a locally finite, open covering of $\mathcal{M}$;

ii) $g_i \in \mathcal{F}(\mathcal{M})$, $g_i(m) \geq 0$ for all $m \in \mathcal{M}$, and $\operatorname{supp}(g_i) \subset V_i$;

iii) $\sum_i g_i(m) = 1$ for all $m \in \mathcal{M}$.

*The partition of unity $\{(V_i, g_i)\}$ is subordinate to the atlas $\mathcal{A}$ if every V_i there is a chart such that $V_i \subset U_{j(i)}$. If any atlas admits a partition of unity, then $\mathcal{M}$ **admits a partition of unity**.*

The sum in Def. 2.120, iii) is well defined since the covering is locally finite. We will not discuss the existence of a partition of unity here but assume it is given for the settings of relevance to us.

---

[105]Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*, Example 6.2.19.C.

Definition 2.121. *Let $\mathcal{M}$ be a manifold with atlas $\mathcal{A} = \{(U_j, \varphi_j)\}$ and $\varphi_j : U_j \to \bar{U}_j \subset \mathbf{E}$ where $\mathbf{E}$ is the model space of $\mathcal{M}$, and let $\{(V_i, g_i)\}$ be a partition of unity subordinate to $\mathcal{A}$. Then for a collection of (r, s) tensor fields $t_j \in \mathcal{T}_s^r(\bar{U}_j)$ that is defined for every chart (U_j, φ_j) there is an (r, s) tensor field $t \in \mathcal{T}_s^r(\mathcal{M})$ on $\mathcal{M}$ given by*

$$t(m) = \sum_i g_i(m) \left(\varphi_j(i)^* t_j(i)\right)(m).$$

The importance of the partition of unity in the above patching is that it ensures that t is smooth and hence forms a tensor field on $\mathcal{M}$. It can be shown that the above construction is well defined, although it is not unique.

Riemannian Metrics and Metrical Operations So far all of our constructions were independent of a metric, and this will also be true for differential forms, the exterior derivative, and integration that are introduced in the next sections. However, a metric will enable us to connect these concepts to ones from classical calculus, and help to gain some intuition for the modern formulation. We will therefore consider metrics in more detail next.

Definition 2.122. *Let $\mathcal{M}$ be a manifold. A **Riemannian metric** is a symmetric, covariant $(0, 2)$ tensor field $g \in \mathcal{T}_2^0(\mathcal{M})$ on $\mathcal{M}$ that is weakly non-degenerate and satisfied $g(m)(v_m, v_m) > 0$ for all $v_m \in T_m\mathcal{M}$ with $v_m \neq 0$. If the Riemannian g is also strongly non-degenerate then it is a **strong Riemannian metric**. When strict positivity $g(m)(v_m, v_m) > 0$ is not satisfied then g is a weak or strong **pseudo-Riemannian metric**. A manifold $\mathcal{M}$ together with a metric is a **Riemannian manifold** $(\mathcal{M}, g)$.*

Recall from Def. 2.29 that a pairing $(\cdot, \cdot) : \mathbf{E} \times \mathbf{E} \to \mathbb{R}$ is weakly non-degenerate if $(v, u) = 0$ for all $u \in \mathbf{E}$ implies $v = 0$, and it is strongly non-degenerate if $v \to (v, \cdot)$ is an isomorphism from $\mathbf{E}$ to to the dual space $\mathbf{E}^*$. Also note that for finite dimensional manifolds the notion of weak and strong Riemannian metrics coincide, that is every weak pairing is also strong.

Example 2.56. The inner product $\langle \cdot, \cdot \rangle$ of a Hilbert space $(\mathcal{H}, \langle , \rangle)$ is a strong Riemannian metric.

Remark 2.76. One often writes $g(m)(v_m, w_m) = \langle v_m, w_m \rangle_m$, recovering the standard notation for Hilbert spaces.

Remark 2.77. Unless noted otherwise, a metric will in the following always refer to a strong Riemannian metric.

Remark 2.78. Let $\mathcal{M}$ be a finite dimensional manifold, and let local bases for $T\mathcal{M}$ and $T^*\mathcal{M}$ be given by $\{\partial/\partial x^1, \ldots, \partial/\partial x^n\}$ and $\{dx^1, \ldots, dx^n\}$. The coordinate expression for a metric is then

$$g(m) = g_{ij}(m) \, dx^i \otimes dx^j$$

and the coefficients are

$$g_{ij}(m) = g(m) \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right)$$

and they are symmetric so that $g_{ij} = g_{ji}$. In matrix form one hence has

$$g(m) = \begin{bmatrix} g_{11} & \cdots & g_{1n} \\ \ldots & \ddots & \ldots \\ g_{1n} & \cdots & g_{nn} \end{bmatrix}$$

for every $m \in \mathcal{M}$ and the matrix coefficients $g_{ij} = g_{ij}(m)$ are smoothly varying functions of the location m. The pairing is then $\bar{v}_m^T g(m) \bar{u}_m$ where $\bar{v}_m$ and $\bar{u}_m$ are the coefficient (column) vectors of $v_m, u_m \in T_m\mathcal{M}$ with respect to $\{\partial/\partial x^1, \ldots, \partial/\partial x^n\}$.

Example 2.57. The standard metric on $\mathbb{R}^3$ is in matrix form

$$g = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

and the expression immediately generalizes to $\mathbb{R}^n$.

We will now generalize raising and lowering indices that was discussed in Remark 2.65 to manifolds.

Definition 2.123. *Let $(\mathcal{M}, g)$ be a Riemannian manifold. Then the **flat map** $\flat : \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}^*(\mathcal{M})$ is defined fiber-wise by*

$$X \to X^\flat \; : \; g_m(X_m, Y_m) = X_m^\flat(Y_m)$$

*for $X, Y \in \mathfrak{X}(\mathcal{M})$. The **sharp map** $\sharp : \mathfrak{X}^* \to \mathfrak{X}(\mathcal{M})$ is*

$$\alpha_m \to X_m \; : \; g(X_m, Y_m) = \alpha_m(Y_m)$$

*for all vector fields $Y \in \mathfrak{X}(\mathcal{M})$. The above maps are known as **musical isomorphism**.*

The musical isomorphisms enable to identify the tangent and the co-tangent space and associate a vector with a co-vector and vice versa.

Remark 2.79. In finite dimensions the musical isomorphisms are given by raising and lowering indices as discussed in Remark 2.65.

Example 2.58. Let $\mathcal{M} = \mathbb{R}^3$ with the standard metric and coordinate system. Then every vector field $X = X^i e_i \in \mathfrak{X}(\mathbb{R}^3)$ has an associated one form field given by $X^\flat = X_i e^1$ where numerically X^i and X_i agree since the metric is diagonally identical, cf. Example 2.57. Analogously, every vector field has an associated 1-form field and numerically the components of both agree.

Example 2.59. Let $(\mathcal{M}, g)$ be a Riemannian manifold. The **gradient** is the vector $(\mathrm{d}f)^\sharp$ associated with the differential $\mathrm{d}f$ of a function $f \in \mathcal{F}(\mathcal{M})$.

2.3.2.4 Differential Forms and Exterior Calculus

Differential forms, antisymmetric and covariant tensors, and the fabric into which they are woven, Cartan's exterior calculus,[106] are central to modern differential geometry and the multitude of fields where geometric and topological questions arise. The importance of the exterior complex arises from its intrinsic coordinate invariance, that results from the tensorial and anti-symmetric nature of differential forms, and its behaviour under the pullback, that enables well defined mappings even when a map is only continuous differentiable. This makes the exterior complex an indispensable tools to study

differential equations, including partial differential ones, which can all be expressed using the exterior derivative,[107] the central operation of the exterior calculus;

topology, where differential forms can be employed to study equivalence classes of manifolds that can be related by continuous differentiable maps;

integration, integrands are differential forms.

Pragmatically, differential forms and exterior calculus can be seen as generalizations of vector calculus in $\mathbb{R}^3$, operations such as grad, div, and curl and theorems such as those by Green, Stokes, and Gauss, to manifolds. Even in Euclidean space, exterior calculus provides much simplification and unification, and we will discuss the reformulation of vector calculus using differential forms throughout the section.

Differential Forms on Linear Spaces[108] As usual, we will introduce differential forms first on linear spaces, and then generalize the concept to manifolds by considering the space of all fibers. Recall from Def. 2.103 that a tensor is anti-symmetric if interchanging any two of its arguments changes its sign; see also Remark 2.62 on permutations.

Definition 2.124. *Let* $\mathbf{E}$ *be a Banach space and* $L_a^k(\mathbf{E}, \mathbb{R}) : \mathbf{E} \times \ldots \times \mathbf{E} \to \mathbb{R}$ *the space of anti-symmetric, k-linear maps on* $\mathbf{E}$. *An* ***exterior k-form*** *is an element in the space* $\bigwedge^k(\mathbf{E})$ *with*

$$\bigwedge^0(\mathbf{E}) = \mathbb{R}$$
$$\bigwedge^1(\mathbf{E}) = E^*$$
$$\bigwedge^k(\mathbf{E}) = L_a^k(\mathbf{E}, \mathbb{R}) \ , \ k > 1.$$

---

[106]For an early historical account see (Chern and Chevalley, "Élie Cartan and his Mathematical Work").

[107]Sharpe, *Differential geometry: Cartan's generalization of Klein's Erlangen program*.

[108]Different authors use different sign and constant conventions for some of the definitions in this section. We follow, as usual, the book by Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*.

For $k = 0$ and $k = 1$ there is no notion of anti-symmetry but the utility of including $\bigwedge^0(\mathbf{E})$ and $\bigwedge^1(\mathbf{E})$ will be apparent when we discuss exterior derivative. The definition of $\mathbf{F}$-valued exterior forms, where $\mathbf{F}$ is a Banach space, is possible using $\mathbf{F}$-valued k-linear maps $L_a^k(\mathbf{E}, \mathbf{F})$. However, we will restrict ourselves to $\mathcal{F} = \mathbb{R}$. Elementary properties of the space $\bigwedge^k(\mathbf{E})$ of k-forms are summarized in the following proposition.

Proposition 2.53. *Let $\mathbf{E}$ be an n-dimensional Banach space with basis $\{e_1, \ldots e_n\}$ and dual basis $\{e^1, \ldots, e^n\}$. Then*

i) $\bigwedge^k(\mathbf{E}) = \{0\}$ *for $k > n$;*

ii) $\dim\left(\bigwedge^k(\mathbf{E})\right) = \frac{n!}{(n-k)!\,k!}$ *;*

iii) for $\alpha \in \bigwedge^k(\mathbf{E})$ and vectors $v_1, \ldots v_k \in \mathbf{E}$ that are linearly dependent, $\alpha(v_1, \ldots, v_k) = 0$.

Definition 2.125. *The **alternation mapping** $\mathbf{A} : T_k^0(\mathbf{E}) \to \bigwedge^k(\mathbf{E})$ takes an arbitrary covariant tensor $t \in T_k^0(\mathbf{E})$ into an exterior form $\mathbf{A}\,t \in \bigwedge^k(\mathbf{E})$ by*

$$(\mathbf{A}\,t)(v_1, \ldots, v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma)\, t(v_{\sigma(1)}, \ldots, v_{\sigma(k)})$$

for $v_i \in \mathbf{E}$, with the summation being over all $k!$ possible permutations.

The alternation mapping is linear and continuous and satisfies $\mathbf{A} \circ \mathbf{A} = \mathbf{A}$. It can hence be considered as a projection operator from T_k^0 onto $\bigwedge^k$.

Remark 2.80. Let $\mathbf{E}$ be a finite dimensional Banach space and $\{e_1, \ldots, e_n\}$ be a basis for $\mathbf{E}$. The coordinate is expression for the alternating map is then

$$(\mathbf{A}\,t)_{i_1, \cdots, i_k} = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}(\sigma)\, t_{\sigma(i_1), \cdots, \sigma(i_k)}. \tag{2.123}$$

Example 2.60. For a $(0, 2)$ tensor $t \in T_s^0(\mathbf{E})$ with coordinates t_{i_1, i_2} the set of all permutations is

$$\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

and the components of the image of the alternation map are hence

$$(\mathbf{A}\,t)_{i_1, i_2} = \frac{1}{2}\left(t_{i_1, i_2} - t_{i_2, i_1}\right).$$

which can be arranged in the form of an anti-symmetric matrix whose diagonal vanishes by Proposition 2.53, iii). For $\mathbf{E} = \mathbb{R}^2$ the only non-trivial components of $\mathbf{A}\,t$ are

$$(\mathbf{A}t)_{12} = \frac{1}{2}\left(t_{12} - t_{21}\right) \qquad (\mathbf{A}t)_{21} = \frac{1}{2}\left(t_{21} - t_{12}\right)$$

which has the form of the usual commutator. Since $(\mathbf{A}t)_{12} = -(\mathbf{A}t)_{21}$ one has in matrix form

$$\mathbf{A}\,t = \left[\begin{array}{cc} 0 & \frac{1}{2}\left(t_{12} - t_{21}\right) \\ -\frac{1}{2}\left(t_{12} - t_{21}\right) & 0 \end{array}\right]$$

which is clearly anti-symmetric.

Definition 2.126. *The **wedge product** $\wedge : T_k^0(\mathbf{E}) \times T_l^0(\mathbf{E}) \to \bigwedge_{k+l}^0(\mathbf{E})$ of two covariant tensors $\alpha \in T_k^0(\mathbf{E})$ and $\beta \in T_l^0(\mathbf{E})$ is*

$$\alpha \wedge \beta = \frac{(k+l)!}{k!\,l!}\,\mathbf{A}(\alpha \otimes \beta).$$

Remark 2.81. The above definition also applies when α and β are exterior forms, and usually we will encounter this case in the following.

Remark 2.82. Wedge products are most conveniently computed using shuffles, permutations $\vec{\sigma}_k^l \in \vec{S}_{k+l}^k$ that are monotonic in each subset $\{1, \ldots k\}$ and $\{k+1, \ldots, l\}$ such that $\vec{\sigma}_k^l(1) < \ldots < \vec{\sigma}_k^l(k)$ and $\vec{\sigma}_k^l(k+1) < \ldots < \vec{\sigma}_k^l(l)$. Using shuffles, the wedge product of $\alpha \in T_k^0(\mathbf{E})$ and $\beta \in T_l^0(\mathbf{E})$ is computed as

$$(\alpha \wedge \beta)(v_1, \ldots, v_{k+l}) = \frac{(k+l)!}{k!\,l!} \sum_{\vec{\sigma} \in \vec{S}_{k+l}^k} \alpha(v_{\bar{\sigma}(1)}, \ldots, v_{\bar{\sigma}(k)})\,\beta(v_{\bar{\sigma}(k+1)}, \ldots, v_{\bar{\sigma}(l)}).$$

where $v_1, \ldots, v_{k+l} \in \mathbf{E}$ are arbitrary.

Remark 2.83. Let $\mathbf{E}$ be finite dimensional. The coordinate expression for the wedge product of $\alpha \in T_k^0$ and $\beta \in T_l^0$ is then

$$(\alpha \wedge \beta)_{i_1, \cdots i_k, i_{k+1}, \cdots, i_{k+l}} = \sum_{\vec{\sigma} \in \vec{S}_{k+l}^k} \alpha_{\vec{\sigma}(i_1), \cdots, \vec{\sigma}(i_k)}\,\beta_{\vec{\sigma}(k+1), \cdots, \vec{\sigma}(k+l)} \qquad (2.124)$$

with summation over all (k, l) shuffles in S_{k+l}^k.

Example 2.61. Let $\alpha, \beta \in \bigwedge^1(\mathbf{E})$ be exterior 1-forms so that $\alpha \wedge \beta \in \bigwedge^2(\mathbf{E})$ is an exterior 2-form. The set of all shuffles $\vec{S}_2^1 = \{\vec{\sigma}^1, \vec{\sigma}^2\}$ of $\{1, 2\}$ is

$$\vec{\sigma}^1 = \left(\begin{array}{c|c} 1 & 2 \\ 1 & 2 \end{array}\right) \quad \vec{\sigma}^2 = \left(\begin{array}{c|c} 1 & 2 \\ 2 & 1 \end{array}\right) \qquad (2.125)$$

which coincides with the set of all permutations, since monotonicity is trivially satisfied in each subset of indices. The wedge product $\alpha \wedge \beta$ for arbitrary $v_1, v_2 \in \mathbf{E}$ is thus

$$(\alpha \wedge \beta)(v_1, v_2) = \alpha(v_{\vec{\sigma}_1^1})\beta(v_{\vec{\sigma}_2^1}) - \alpha(v_{\vec{\sigma}_1^2})\beta(v_{\vec{\sigma}_2^2}) \qquad (2.126\text{a})$$

$$= \alpha(v_1)\beta(v_2) - \alpha(v_2)\beta(v_1), \qquad (2.126\text{b})$$

which takes the form of the commutator of two covectors.

Example 2.62. Let $\alpha \in \bigwedge^2(\mathbf{E})$ be an exterior 2-form and $\beta \in \bigwedge^1(\mathbf{E})$ be an exterior 1-form so that $\alpha \wedge \beta \in \bigwedge^3(\mathbf{E})$ is an exterior 3-form. The set of all shuffles $\vec{S}_3^2 = \{\vec{\sigma}^1, \vec{\sigma}^2, \vec{\sigma}^3\}$ of $\{1, 2, 3\}$ is

$$\vec{\sigma}^1 = \begin{pmatrix} 1 & 2 & \vdots & 3 \\ 1 & 2 & \vdots & 3 \end{pmatrix} \quad \vec{\sigma}^2 = \begin{pmatrix} 1 & 2 & \vdots & 3 \\ 1 & 3 & \vdots & 2 \end{pmatrix} \quad \vec{\sigma}^3 = \begin{pmatrix} 1 & 2 & \vdots & 3 \\ 2 & 3 & \vdots & 1 \end{pmatrix}, \quad (2.127)$$

with monotonicity in the subsets of indices. The wedge product $\alpha \wedge \beta$ for arbitrary $v_1, v_2, v_2 \in \mathbf{E}$ is thus

$$(\alpha \wedge \beta)(v_1, v_2, v_3) \quad = \alpha(v_1, v_2)\beta(v_3) - \alpha(v_1, v_3)\beta(v_2) + \alpha(v_2, v_3)\beta(v_1).$$
$$(2.128)$$

The properties of the wedge product are summarized in the next proposition.

Proposition 2.54. *Let $\alpha \in T_k^0$, $\beta \in T_l^0$, and $\gamma \in T_m^0$. Then*

i) $\alpha \wedge \beta = \mathbf{A}\alpha \wedge \beta = \alpha \wedge \mathbf{A}\beta$;

ii) $\wedge$ is bilinear;

iii) $\alpha \wedge \beta = (-1)^{kl}\beta \wedge \alpha$;

iv) $\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma = \frac{(k+l+m)!}{k!\, l!\, m!} \mathbf{A} (\alpha \otimes \beta \otimes \gamma)$.

Example 2.63. It follows from Theorem 2.54, iii) that if $\alpha \in \bigwedge^1$ is a 1-form then $\alpha \wedge \alpha = 0$.

Definition 2.127. *The **Grassmann** or **exterior algebra** $\bigwedge(\mathbf{E})$ of a Banach space $\mathbf{E}$ is the direct sum $\bigwedge(\mathbf{E}) = \bigwedge^0(\mathbf{E}) \oplus \bigwedge^1(\mathbf{E}) \oplus \ldots$ together with its vector space structure and the algebra product given by the wedge product.*

Remark 2.84. The exterior algebra is a graded algebra, that is the product of an element in the kth grade and in the lth grade is an element in the $(k + l)^{\text{th}}$ grade. Other examples of graded algebras are polynomials and the Fourier basis under the usual point-wise multiplication of functions. Hence, analogously to polynomials that are usually written as linear combination of monomials, an element $\alpha \in \bigwedge(\mathbf{E})$ can be written as

$$\alpha = a_1 \, \alpha^1 + a_2 \, \alpha^2 + \ldots$$

where the $a_i \in \mathbb{R}$ and $\alpha_i \in \bigwedge^k(\mathbf{E})$.

Properties of the exterior algebra are summarized in the following proposition.

Proposition 2.55. *Let $\mathbf{E}$ be an n-dimensional Banach space with basis $\{e_1, \ldots e_n\}$ and dual basis $\{e^1, \ldots, e^n\}$. Then $\dim(\bigwedge(\mathbf{E})) = 2^n$ and a basis for $\bigwedge^k(\mathbf{E})$ is given by $\{e^{i_1} \wedge \ldots \wedge e^{i_k} \mid 1 \leq i_1 < \ldots < i_k \leq n\}$ for all possible $\{i_1, \ldots, i_k\}$.*

Example 2.64. Let $\mathbf{E} = \mathbb{R}^3$ with the canonical basis $\{e_1, e_2, e_3\}$. Then

$$\dim\left(\bigwedge^1(\mathbb{R}^3)\right) = \frac{3!}{(3-1)!\,1!} = \frac{6}{2} = 3 \tag{2.129a}$$

$$\dim\left(\bigwedge^2(\mathbb{R}^3)\right) = \frac{3!}{(3-2)!\,2!} = \frac{6}{2} = 3 \tag{2.129b}$$

$$\dim\left(\bigwedge^3(\mathbb{R}^3)\right) = \frac{3!}{(3-3)!\,3!} = \frac{6}{6} = 1 \tag{2.129c}$$

and bases for the grades are given by

$$\bigwedge^1(\mathbb{R}^3) : \{e^1, e^2, e^3\} \tag{2.130a}$$

$$\bigwedge^2(\mathbb{R}^3) : \{e^2 \wedge e^3, e^1 \wedge e^3, e^1 \wedge e^2\} \tag{2.130b}$$

$$\bigwedge^1(\mathbb{R}^3) : \{e^1 \wedge e^2 \wedge e^3\}. \tag{2.130c}$$

Let $\alpha, \beta \in \bigwedge^1(\mathbb{R}^3)$ with coordinate expression $\alpha = \alpha_i e^i$ and $\beta = \beta_i e^i$. Then by Eq. 2.124 and Example 2.62 the basis representation of the wedge product $\alpha \wedge \beta$ is

$$\alpha \wedge \beta = (\alpha_2\beta_3 - \alpha_3\beta_2)(e^2 \wedge e^3)$$
$$+ (\alpha_3\beta_1 - \alpha_1\beta_3)(e^3 \wedge e^1) + (\alpha_1\beta_2 - \alpha_2\beta_1)(e^1 \wedge e^2) \tag{2.131}$$

and the basis function coefficients for the basis for $\bigwedge^2(\mathbb{R}^3)$ in Eq. 2.130b are

$$(\alpha \wedge \beta)_{23} = \alpha_2\beta_3 - \alpha_3\beta_2$$
$$(\alpha \wedge \beta)_{13} = \alpha_3\beta_1 - \alpha_1\beta_3$$
$$(\alpha \wedge \beta)_{12} = \alpha_1\beta_2 - \alpha_2\beta_1.$$

The coordinate expression for the wedge product of two 1-forms in $\mathbb{R}^3$ is similar to the cross product, and we will in fact be able to recover the usual cross product in the next section after we introduced the Hodge dual.

Remark 2.85. For an n-dimensional vector space $\mathbf{E}$ the top grade is one dimensional, that is $\dim\left(\bigwedge^n(\mathbf{E})\right) = 1$ with a basis given by $\{e^1 \wedge \ldots \wedge e^n\}$. Hence, the space of n-forms is often identified with the space of functions on $\mathbf{E}$. However, as we will see in the sequel, functions and n-forms behave different under a change of coordinates and it is hence important to distinguish them.

The following proposition characterizes a generalized notion of orthogonality for exterior forms.

Proposition 2.56. *Let $\alpha \in \bigwedge^1(\mathbf{E})$ and $\gamma \in \bigwedge^k(\mathbf{E})$. Then $\alpha \wedge \gamma = 0$ if and only if there exists a $\beta \in \bigwedge^{k-1}(\mathbf{E})$ such that $\gamma = \alpha \wedge \beta$.*
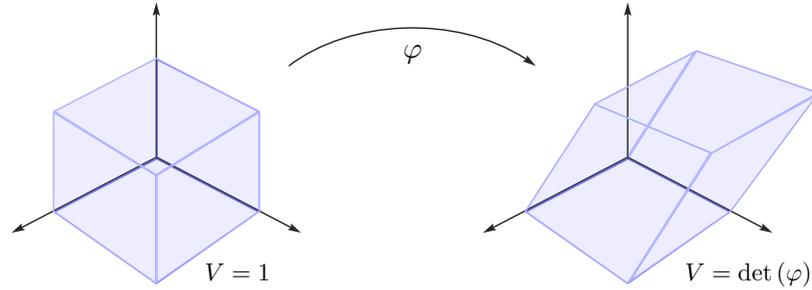
Figure 2.34: For a mapping $\varphi : \mathbb{R}^3 \to \mathbb{R}^3$ the determinant $\det(\varphi(x))$ measures the change in the volume of the unit cube.

To provide some additional intuition for exterior forms on linear space we will in the following show how these generalize the determinant and volume known from linear algebra and vector calculus. Let us begin by recalling the definition of the determinant for $\mathbb{R}^3$, see Fig. 2.34.

Definition 2.128. *Let $\varphi : \mathbb{R}^3 \to \mathbb{R}^3$. Then the determinant $\det(\varphi(x))$ is the volume of the image of the unit cube at x under φ.*

To define the determinant using exterior forms and for arbitrary vector spaces we will require the pullback of forms. Since forms are special covariant tensors, Def. 2.109 and Proposition 2.48 apply, and we summarize the properties in the following proposition.

Proposition 2.57. *Let $\mathbf{E}, \mathcal{F}$ be normed vector spaces, and let $\varphi : \mathbf{E} \to \mathcal{F}$ be a map between the spaces. Then for $\alpha \in \bigwedge^k(\mathbf{F})$ and $\beta \in \bigwedge^l(\mathbf{F})$ the pullback $\varphi^* : \bigwedge^k(\mathcal{F}) \to \bigwedge^k(E)$ commutes with the wedge product such that*

$$\varphi^*(\alpha \wedge \beta) = \varphi^*\alpha \wedge \varphi^*\beta$$

and it is closed in that

$$\varphi^*\left(\bigwedge{}^k(\mathbf{F})\right) \subset \bigwedge^k(\mathbf{E}).$$

Using the pullback for exterior forms we can now generalize the determinant.

Definition 2.129. *Let $\mathbf{E}$ be a finite dimensional Banach space with $\dim(\mathbf{E}) = n$ and $\varphi \in L(\mathbf{E}, \mathbf{E})$. Then the **determinant** is the unique constant $\det(\varphi) \in \mathbb{R}$ such that $\varphi^* : \bigwedge^n(\mathbf{E}) \to \bigwedge^n(\mathbf{E})$ for all $\omega \in \bigwedge^n(\mathbf{E})$ satisfies*

$$\varphi^*\omega = \det(\varphi)\,\omega.$$

The classical definition of the determinant employs a specific representation for the map $\varphi \in L(\mathbf{E}, \mathbf{E})$ whereas Def. 2.86 is intrinsically coordinate independent by its tensorial nature.

Remark 2.86. Let $\mathbf{E}$ be finite dimensional with basis $\{e_1, \ldots, e_n\}$ and $\varphi \in L(\mathbf{E}, \mathbf{E})$ have the matrix representation A_i^j such that $\varphi(e_i) = A_i^j \, e_i$. Then for $\omega = e^1 \wedge \ldots \wedge e^n \in \bigwedge^n(\mathbf{E})$ we have by the definition of the pullback, Def. 2.107 and Def. 2.108,

$$\varphi^* \omega = \varphi^*(e^1 \wedge \ldots \wedge e^n)(e_1, \ldots, e_n) = (e^1 \wedge \ldots \wedge e^n)(\varphi(e_1), \ldots, \varphi(e_n)).$$

Repeatedly using Proposition 2.54, iv) gives

$$e^1 \wedge \ldots \wedge e^n = k! \, \mathbf{A}(e^1 \otimes \ldots \otimes e^n)$$

and by linearity and $e^k(\varphi(e_i)) = e^k(A_j^i e_i) = A_i^k$ we have with the definition of the alternation map in Def. 2.125 that

$$
\begin{aligned}
\varphi^* \omega &= \sum_{\sigma \in S_k} \text{sgn}\,(\sigma)\,(e^1 \otimes \ldots \otimes e^n)(\varphi(e_{\sigma(1)}), \ldots, \varphi(e_{\sigma(n)})) \\
&= \sum_{\sigma \in S_k} \text{sgn}\,(\sigma)\, A_1^{\sigma(1)} \ldots A_n^{\sigma(n)} \\
&= \sum_{\sigma \in S_k} \text{sgn}\,(\sigma) \prod_i A_i^{\sigma(i)}
\end{aligned}
$$

which is the Leibniz formula for the determinant, showing that Def. 2.129 agrees with the classical definition in the literature, see Def. 2.128.

Example 2.65. Let $R : \mathbb{R}^3 \to \mathbb{R}^3$ be a rotation. Then $\det\,(R) = 1$ and hence there is no volume distortion under the map R; which states the well known fact that rotations are isometries of Euclidean space.

The usual properties of the determinant are from our point of view corollaries to those of the pullback for the special case of a form of maximal degree, cf. Proposition 2.48.

Corollary 2.10. *Let $\varphi, \psi \in L(\mathbf{E}, \mathbf{E})$ and $\mathbf{E}$ be finite dimensional. Then*

i) $\det\,(\varphi \circ \psi) = \det\,(\varphi)\,\det\,(\psi)$;

ii) if φ is the identity then $\det\,(\varphi) = 1$;

iii) φ is an isomorphism if and only if $\det\,(\varphi) \neq 0$ and in this case $\det\,(\varphi^{-1}) = 1/\det\,(\varphi)$.

Since $\bigwedge^n(\mathbf{E})$ is one dimensional, we can use the two equivalence classes of elements in $\bigwedge^n(\mathbf{E})$ defined by the non-vanishing sign of their coefficient to define an orientation of $\mathbf{E}$.

Definition 2.130. *Let $\mathbf{E}$ be a vector space. The nonzero elements of $\bigwedge^n(\mathbf{E})$ are the **volume forms** or **volume elements** on $\mathbf{E}$. Two volume forms $\omega_1, \omega_2 \in \bigwedge^n(\mathbf{E})$ are equivalent if there is a $c > 0$ such that $\omega_1 = c\,\omega_2$. A so defined equivalence class $[\omega]$ of volume elements is called an **orientation***

on **E**. An **oriented vector space** is the tuple $(\mathbf{E}, [\omega])$ of a vector space **E** together with an orientation, and $-[\omega]$ is the reverse orientation on the space. A basis $\{e_1, \ldots, e_n\}$ is **positively oriented** when $\omega(e_1, \ldots e_n) > 0$ for $\omega \in [\omega]$; otherwise it is **negatively oriented**.

An orientation on a vector space induces an equivalence class of ordered bases and using a reference basis the orientations can also be defined using the determinant since any change of coordinates $\varphi \in L(\mathbf{E}, \mathbf{E})$ that preserves the orientation has a positive determinant.

The above definition of a volume element is independent of any additional structure such as a metric, although we know that in $\mathbb{R}^3$ the volume is related to the notion of length and area, all of which are consistent; for example the volume of a parallelepiped is its base area times its height. A volume form that is consistent with an additional structure on the space such as a metric leads to the notion of a density.

Definition 2.131. *Let α be a real number and* **E** *an n-dimensional vector space. A continuous mapping*

$$\rho : \underbrace{\mathbf{E} \times \ldots \times \mathbf{E}}_{n \ times} \to \mathbb{R}$$

*is called an α-**density** on* **E** *if*

$$\rho(\varphi(v_1), \ldots, \varphi(v_n)) = |\det(\varphi)|^\alpha \rho(v_1, \ldots, v_n)$$

for all $v_i \in \mathbf{E}$ and $\varphi \in L(\mathbf{E}, \mathbf{E})$. The space of all α-densities is denote by $|\bigwedge|^\alpha(\mathbf{E})$ and when $\alpha = 1$ we will use $|\bigwedge|(\mathbf{E})$ and simply call them densities.

Remark 2.87. The concept of densities can be generalized to pseudo-forms that do *not* change sign when the orientation of the underlying space is reversed.[109]

Remark 2.88. An α-density $|\omega|^\alpha \in |\bigwedge|^\alpha$ can be constructed from a volume element $\omega \in \bigwedge^n$ as

$$|\omega|^\alpha(e_1, \ldots, e_n) = |\omega(e_1, \ldots, e_n)|^\alpha.$$

Remark 2.89. The difference between volume forms and α-densities becomes significant when integration is considered, in particular on non-orientable manifolds. For example, for a function $f : \mathbb{R} \to \mathbb{R}$ the integral is

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx.$$

The change in the sign of the integral under a reversal of the orientation of $[a, b]$ indicates that dx is a volume form; by definition, the sign is invariant under a change of orientation for densities. Note that the above integral would also *not* change sign when determined using measure theory. Densities can thus be considered to be an intermediary between exterior forms and measures.

---

[109]See the book by Frankel (*The Geometry of Physics*) for details.

Definition 2.132. *Let* $\mathbf{E}$ *be an* n-*dimensional vector space and* $g \in T_2^0(\mathbf{E})$ *be a metric on* $\mathbf{E}$. *A basis* $\{e_1, \ldots, e_n\}$ *is called* g-***orthonormal*** *if* $g(e_i, e_j) = \pm\delta_{ij}$. *The* ***signature*** $s(g)$ *of* g *is* $s(g) = \sum_i g(e_i, e_i)$ *and the* ***index*** $\mathrm{ind}\,(g)$ *is the number of basis vectors* e_i *such that* $g(e_i, e_i) = -1$.

Example 2.66. The flat Minkowski metric is given in matrix form by

$$
g_M = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
$$

For the signature and index of g_M we hence have $s(g_M) = 3$ and $\mathrm{ind}\,(g_M) = -1$.

Definition 2.133. *Let* $\mathbf{E}$ *be an* n-*dimensional vector space and* $g \in T_2^0(\mathbf{E})$ *be a metric on* $\mathbf{E}$, *and* $[\omega]$ *be an orientation of* $\mathbf{E}$.

   i) *The* g-***volume*** *defined by the metric* g *is the unique volume element* $\mu_g \in [\omega]$ *such that* $\mu(e_1, \ldots, e_n) = 1$ *for any* g-*orthonormal basis* $\{e_1, \ldots e_n\}$. *Moreover* $\mu = e^1 \wedge \ldots \wedge e^n$ *where* $\{e^1, \ldots, e^n\}$ *is the dual basis of* $\{e_1, \ldots e_n\}$. *For an arbitrary, not necessarily* g-*orthonormal basis* $\{\bar{e}_1, \ldots, \bar{e}_n\}$ *with dual basis* $\{\bar{e}^1, \ldots, \bar{e}^n\}$ *the* g-*volume is given by*

$$
\mu = \sqrt{|g(\bar{e}_i, \bar{e}_j)|}\, \bar{e}^1 \wedge \ldots \wedge \bar{e}^n.
$$

   ii) *A* g-α-***density*** $|\mu|^\alpha$ *is a density such that* $|\mu|^\alpha(e_1, \ldots, e_n) = 1$ *for all* g-*orthonormal bases* $\{e_1, \ldots, e_n\}$ *and it is given in terms of the dual basis* $\{e^1, \ldots, e^n\}$ *by* $|\mu|^\alpha = |e^1 \wedge \ldots \wedge e^n|^\alpha$.

The above theorem defines a preferred volume element and α-density consistent with the metric on a space where such additional structure is available. A metric does not only induce a pairing on vectors but also on exterior k-forms, as we will see next

Definition 2.134. *Let* $\mathbf{E}$ *be an* n-*dimensional vector space and* $g \in T_2^0(\mathbf{E})$ *be a metric on* $\mathbf{E}$. *The* ***metric pairing*** $\langle\!\langle\,,\rangle\!\rangle : \bigwedge^k(\mathbf{E}) \times \bigwedge^k(\mathbf{E}) \to \mathbb{R}$ *between two* k-*forms* $\alpha = \alpha_{i_1, \cdots, i_k} e^{i_1} \wedge \ldots \wedge e^{i_k} \in \bigwedge^k(\mathbf{E})$ *and* $\beta = \beta_{i_1, \cdots, i_k} e^{i_1} \wedge \ldots \wedge e^{i_k} \in \bigwedge^k(\mathbf{E})$ *is*

$$
\langle\!\langle \alpha, \beta \rangle\!\rangle = \sum_{i_1 < \cdots < i_k} \alpha_{i_1, \cdots, i_k} \beta^{i_1, \cdots, i_k} = \sum_{i_1 < \cdots < i_k} \alpha_{i_1, \cdots i_k} \left(g^{j_1, i_1} \ldots g^{j_k, i_k} \beta_{j_1, \cdots j_k} \right).
$$

With the above pairing we can introduce the important concept of the Hodge dual.

Definition 2.135. *Let* $\mathbf{E}$ *be an* n-*dimensional, oriented vector space and* $g \in T_2^0(\mathbf{E})$ *be a metric on* $\mathbf{E}$, *and* μ *be the* g-*volume on* $\mathbf{E}$. *The* ***Hodge dual*** $\star : \bigwedge^k(\mathbf{E}) \to \bigwedge^{n-k}(\mathbf{E})$ *is the unique isomorphism between* $\bigwedge^k(\mathbf{E})$ *and* $\bigwedge^{n-k}(\mathbf{E})$ *such that for* $\alpha, \beta \in \bigwedge^k(\mathbf{E})$

$$
\alpha \wedge \star\beta = \langle\!\langle \alpha, \beta \rangle\!\rangle \mu.
$$

It is important to keep in mind that the Hodge dual is a metric-dependent operation, since $\langle\!\langle\,,\rangle\!\rangle$ is defined using the metric g on $\mathbf{E}$.

Example 2.67. Let $\mathbf{E}$ be $\mathbb{R}^3$ with the canonical basis $\{e_1, e_2, e_3\}$. Then the dual basis $\{e^1, e^2, e^3\}$ provides a basis for $\bigwedge^1(\mathbf{E}) \cong \mathbf{E}^*$ and

$$\star e^1 = e^2 \wedge e^3 \qquad \star e^2 = -e^1 \wedge e^3 \qquad \star e^3 = e^1 \wedge e^3$$

hence provides a basis for $\bigwedge^2(\mathbf{E})$, cf. Example 2.64.

We summarize important properties of the Hodge dual in the following proposition.

Proposition 2.58. *Let* $\mathbf{E}$ *be an n-dimensional, oriented vector space with metric* $g \in T_2^0(\mathbf{E})$*, and* μ *the induced g-volume on* $\mathbf{E}$*. For* $\alpha, \beta \in \bigwedge^k(\mathbf{E})$ *then*

i) $\alpha \wedge \star\beta = \beta \wedge \star\alpha = \langle\!\langle \alpha, \beta \rangle\!\rangle \mu$;

ii) for 1 *being the multiplicative identity on* $\mathbf{E}$*,* $\star 1 = \mu$ *and* $\star\mu = \pm 1$;

iii) $\star\star\alpha = \pm\alpha$;

iv) $\langle\!\langle \alpha, \beta \rangle\!\rangle = \pm\langle\!\langle \alpha, \beta \rangle\!\rangle$;

and where the sign depends on the signature of the metric.

Remark 2.90. An expression for the Hodge dual of a basis function of an exterior form that is often useful in practice is

$$\star(e^{i_1} \wedge \ldots \wedge e^{i_k})$$

$$= \sqrt{|\det(g_{ij})|} \sum \mathrm{sgn} \begin{pmatrix} 1 \cdots n \\ j_1 \cdots j_n \end{pmatrix} g^{i_1, j_1} \ldots g^{i_k, j_k} \, e^{j_{k+1}} \wedge \ldots \wedge e^n.$$

with summation over all $(k, n-k)$ shuffles.

Example 2.68. In Def. 2.123 we introduced the musical isomorphisms. Together with the Hodge dual these enable to write the cross product in $\mathbb{R}^3$ as

$$u \times v = (\star(u^\flat \wedge v^\flat))^\sharp \in \mathbb{R}^3$$

for $u, v \in \mathbb{R}^3$. In fact, in coordinates we have $u = u^i e_1$ and $v = v^i e_i$. Hence, $u^\flat = u_i e^i$ and $v^\flat = v_i e^i$ where numerically $v_i = v^i$ and $u_i = u^i$ since the standard metric on $\mathbb{R}^3$ is diagonally identical. From Example 2.64 we know

$$u^\flat \wedge v^\flat =$$

$$(u_2 v_3 - u_3 v_2)(e^2 \wedge e^3) + (u_3 v_1 - u_1 v_3)(e^3 \wedge e^1) + (u_1 v_2 - u_2 v_1)(e^1 \wedge e^2).$$
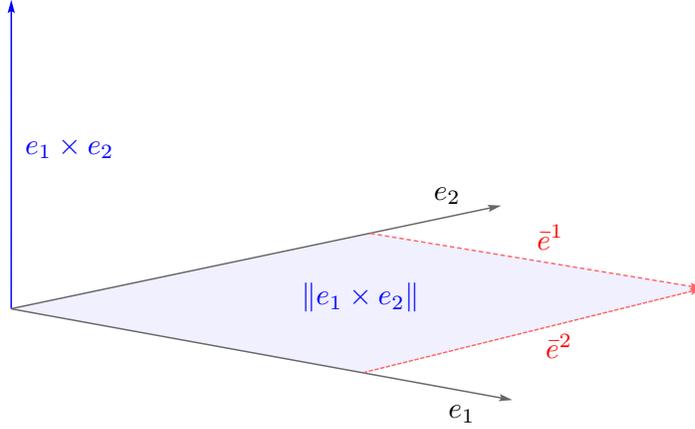
Figure 2.35: The cross product determines the area spanned by two vectors in a plane but it is a vector orthogonal to the plane, hence requiring an embedding. Using the wedge product and exterior forms and intrinsic definition is possible.

Using the expression for the Hodge dual of the basis functions in Example 2.67, Proposition 2.58, and the antisymmetry of exterior forms so that $\alpha \wedge \beta = -\beta \wedge \alpha$ we obtain

$$\star(u^\flat \wedge v^\flat) = (u_2 v_3 - u_3 v_2)e^1 + (u_3 v_1 - u_1 v_3)e^2 + (u_1 v_2 - u_2 v_1)e^3.$$

Finally, applying the musical isomorphism $\sharp$, that again numerically leaves the coordinates intact, yields

$$(\star(u^\flat \wedge v^\flat))^\sharp = (u_2 v_3 - u_3 v_2)e_1 + (u_3 v_1 - u_1 v_3)e_2 + (u_1 v_2 - u_2 v_1)e_3$$

which is the usual expression for the cross product. Although the expression using the wedge product and the musical isomorphisms does not necessarily look simpler, it is an intrinsic expression that enables to generalize the concept to arbitrary coordinate systems and vector spaces, and ultimately to manifolds. Additionally, in many situations where the cross product arises one has intrinsically exteriror or differential forms and their wedge product is the natural operation to be performed. For example, in $\mathbb{R}^3$ the cross product determines the size of the are spanned by two vectors and for instance

$$\|e_1 \times e_2\| = \|e_1\| \, \|e_2\| \sin(\angle(e_1, e_2)) = 1 \tag{2.132}$$

is the area spanned by the canonical basis vectors $e_1, e_2 \in \mathbb{R}^3$, cf. Fig. However, although $e_1 = \bar{e}_1$ and $e_2 = \bar{e}_2$ coincide with the canonical basis vectors $\bar{e}_1, \bar{e}_2 \in \mathbb{R}^2$, and the problem is hence entirely defined in the plane, the cross product $e_1 \times e_2$ is a vector in $\mathbb{R}^3$ orthogonal to the plane spanned e_1 and e_2. Using the wedge product and the exterior forms $\bar{e}^1$ and $\bar{e}^1$ dual to $\bar{e}_1$ and $\bar{e}_2$ the area can be defined by

$$\bar{e}^1 \wedge \bar{e}^2 = 1 \tag{2.133}$$

without the need to employ the embedding of $\mathbb{R}^2$ in $\mathbb{R}^3$. It should be noted, however, that coordinate expression resembling the cross product can arise not only through the wedge product but also the pullback, as we will see in the sequel.

Differential Forms   In the following we will extent exterior forms from linear spaces to vector bundles. As starting point we will use local vector bundles.

Definition 2.136. *Let* $\mathbf{E}, \mathbf{F}, \mathbf{E}', \mathbf{F}'$ *be Banach spaces and* $U \subset \mathbf{E}$ *and* $U' \subset \mathbf{E}'$ *be open. Consider the local vector bundle map* $\varphi : U \times \mathbf{F} \to U' \times \mathbf{F}'$ *that is an isomorphism* $\varphi_u : \mathbf{F}_u \to \mathbf{F}'_{\varphi_0(u)}$ *on each fiber. The induced map*

$$\varphi_* : U \times \bigwedge^k (\mathbf{F}) \to U' \times \bigwedge^k (\mathbf{F}')$$

on the kth *grade of the exterior algebra on each fiber* $\mathbf{F}_u$ *is defined by* $\varphi(u, \alpha) = (\varphi_0(u), \varphi_{u*}\alpha)$.

It can be shown that if $\varphi : U \times \mathbf{F} \to U' \times \mathbf{F}'$ is a local vector bundle map in the sense of Def. 2.76 then $\varphi_* : U \times \bigwedge^k(\mathbf{F}) \to U' \times \bigwedge^k(\mathbf{F}')$ is also a local vector bundle map, and if φ is a local vector bundle isomorphism then so is φ_*. As before, this ensures that φ_* can be employed as transition map between charts.

Definition 2.137. *Let* $\varphi : E \to B$ *be a vector bundle with base* B. *Then*

$$\bigwedge^k(E) = \bigcup_{b \in B} \bigwedge^k(E_b),$$

where E_b *is the fiber over* $b \in B$. *The fiber projection* $\pi_{\bigwedge}^k : \bigwedge^k(E) \to B$ *is defined by* $\pi_{\bigwedge}^k(t) = b$ *when* $t \in \bigwedge^k(E_b)$.

One can show that a vector bundle atlas for $\pi : E \to B$ induces a vector bundle atlas on $\bigwedge^k(E)$ and hence $\bigwedge^k(E)$ is also a vector bundle, which is the expected and desired structure. Specializing Def. 2.137 to the tangent bundle leads to exterior differential forms.

Definition 2.138. *Let* $\tau_{\mathcal{M}} : T\mathcal{M} \to \mathcal{M}$ *be the tangent bundle of a manifold* $\mathcal{M}$. *The vector bundle* $\bigwedge^k(\mathcal{M})$ *of exterior k-forms on* $T\mathcal{M}$ *is* $\bigwedge^k(\mathcal{M}) = \bigwedge^k(T\mathcal{M})$ *with bundle projection* $\pi_{\bigwedge}^k : \bigwedge^k(\mathcal{M}) \to \mathcal{M}$. *The space of sections of* $\bigwedge^k(\mathcal{M})$ *for* $k > 1$ *is the* **space of differential k-forms**

$$\Omega^k(\mathcal{M}) = \Gamma\left(\bigwedge^k(\mathcal{M})\right)$$

and one define $\Omega^0(\mathcal{M}) = \mathcal{F}(\mathcal{M})$ *and* $\Omega^1(\mathcal{M}) = \mathfrak{X}^*(\mathcal{M})$. *The* **exterior algebra** $\Omega(\mathcal{M})$ *on* $\mathcal{M}$ *is the direct sum*

$$\Omega(\mathcal{M}) = \Omega^0(\mathcal{M}) \oplus \Omega^1(\mathcal{M}) \oplus \ldots$$

together with its infinite-dimensional vector space structure and the algebra product given by the wedge product, defined fiber-wise over all of $\mathcal{M}$. The elements $\alpha^k \in \Omega^k(\mathcal{M})$ are known as **exterior differential forms of degree k** *or* **differential k-forms**.

The extension of the wedge product to all of $\mathcal{M}$ is defined fiber-wise by $(\alpha \wedge \beta)(m) = \alpha(m) \wedge \beta(m)$ for $\alpha \in \Omega^k$ and $\beta \in \Omega^l(\mathcal{M})$. As the multiplication in the exterior algebra of differential forms the wedge product is hence

$$\wedge : \Omega^k(\mathcal{M}) \times \Omega^l(\mathcal{M}) \to \Omega^{k+l}(\mathcal{M}) \tag{2.134}$$

which also shows that the exterior algebra is again a graded algebra. For $\alpha^0 = f \in \Omega^k(\mathcal{M})$ the wedge product reduces to the fiber-wise scalar multiplication $f\beta \in \Omega^l(\mathcal{M})$ for any $\beta \in \Omega^l(\mathcal{M})$.

Example 2.69. A 2-form $\omega \in \Omega^2(\mathcal{M})$ assigns to each tangent space $T_m\mathcal{M}$ an anti-symmetric bilinear map $\omega_m : T_m\mathcal{M} \times T_m\mathcal{M} \to \mathbb{R}$.

Remark 2.91. Let $\mathcal{M}$ be a finite dimensional manifold and (U, φ) a chart for $\mathcal{M}$ with local bases $\{\partial/\partial x^1, \ldots, \partial/\partial x^n\}$ and $\{dx^1, \ldots, dx^n\}$ for $T\mathcal{M}$ and $T^*\mathcal{M}$, respectively. Following Def. 2.114, the coordinate expression for a differential k-form $\alpha \in \Omega^k(\mathcal{M})$ at $u \in U \subset \mathcal{M}$ is then

$$\alpha(u) = \alpha_{i_1, \cdots, i_k}(u)\, dx^{i_1} \wedge \ldots \wedge dx^{i_k} \; , \; i_1 < \ldots < i_k \tag{2.135}$$

with summation over all monotonic indices $i_1 < \ldots < i_k$ implied, and the coefficients are given by

$$\alpha_{i_1, \ldots i_k}(u) = \alpha \left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n} \right)(u). \tag{2.136}$$

Remark 2.92. Formally it is possible to consider the addition of differential forms of different degrees and it is an operation often employed in geometric algebra. We will not employ elements in $\Omega(\mathcal{M})$ of mixed degree.

In the following, we will introduce the push-forward and pullback of differential forms between manifolds.

Definition 2.139. *Let $\mathcal{M}, \mathcal{N}$ be manifolds and $\varphi : \mathcal{M} \to \mathcal{N}$ be a smooth map. The* **pullback** $\varphi^* : \Omega^k(\mathcal{N}) \to \Omega^k(\mathcal{M})$ *of $\omega \in \Omega^k(\mathcal{N})$ at $m \in \mathcal{M}$ is*

$$(\varphi^*\omega)(m) = (T_m\varphi)^* \circ \omega \circ \varphi(m)$$

so that for $v_1, \ldots, v_k \in T_m\mathcal{M}$ one has

$$(\varphi^*\omega)_m(v_1, \ldots, v_k) = \omega((T_m\varphi) \cdot v_1, \ldots, (T_m\varphi) \cdot v_k).$$

For $f \in \Omega^0(\mathcal{M}) = \mathcal{F}(\mathcal{M})$ the pullback is defined as $\varphi^* f = f \circ \varphi$.

In the above definition, $T_m\varphi$ is the tangent map relating the tangent space $T_m\mathcal{M}$ to $T_{\varphi(m)}\mathcal{N}$.

Remark 2.93. Notationally, one has to be careful to distinguish the cotangent lift $T^*\varphi : T^*\mathcal{N} \to T^*\mathcal{M}$ and the pullback $(T\varphi)^* : \Omega^k(T\mathcal{N}) \to \Omega^k(T\mathcal{M})$ of the tangent map $T\varphi : T\mathcal{M} \to T\mathcal{N}$, cf. Def. 2.73, which will be introduced in the following.

The properties of the pullback are summarized in the following proposition.

Theorem 2.15. *Let $\varphi : \mathcal{M} \to \mathcal{N}$ and $\psi : \mathcal{N} \to \mathcal{K}$ be smooth mappings between manifolds. Then*

   i) $(\psi \circ \varphi)^* = \varphi^* \circ \psi^*$;

   ii) if $\mathrm{id} : \mathcal{M} \to \mathcal{M}$ is the identity on $\mathcal{M}$ then $\mathrm{id}^* : \Omega^k(\mathcal{M}) \to \Omega^k(\mathcal{M})$ is the identity;

   iii) if φ is a diffeomorphism then φ^* is an isomorphism and $(\varphi^*)^{-1} = (\varphi^{-1})^*$;

   iv) for $\alpha \in \Omega^k(\mathcal{N})$ and $\beta \in \Omega^l(\mathcal{N})$, $\varphi^*(\alpha \wedge \beta) = \varphi^*\alpha \wedge \varphi^*\beta$.

Next we will introduce the exterior derivative, the central operation of exterior calculus. It generalizes the differential of functions, a map from 0-forms to 1-forms, to a map from arbitrary k-forms to $k + 1$-forms. The exterior derivative is naturally defined on the exterior algebra and it can hence be introduced in the form of a theorem.

Theorem 2.16. *Let $\mathcal{M}$ be a manifold and $U \subset \mathcal{M}$ be open. The **exterior derivative** is the unique mapping $\mathrm{d} : \Omega^k(U) \to \Omega^{k+1}(U)$ such that*

   i) d is an anti-derivation of the exterior algebra that is $\mathbb{R}$-linear and satisfies the Leibniz rule, that is for $\alpha \in \Omega^k(U)$ and $\beta \in \Omega^l(U)$,

$$\mathrm{d}(\alpha \wedge \beta) = \mathrm{d}\alpha \wedge \beta + (-1)^{kl}\alpha \wedge \mathrm{d}\beta;$$

   ii) for $f \in \Omega^0(U) = \mathcal{F}(U)$, $\mathrm{d}f$ is the differential of functions;

   iii) $\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} = 0$;

   iv) d is a local operator that is natural with respect to restrictions, that is for $V \subset U \subset \mathcal{M}$ for $\alpha \in \Omega^k(U)$ one has $(\mathrm{d}\alpha)\mid_V = \mathrm{d}(\alpha\mid_V)$.

The restriction to an open neighborhood $U \subset \mathcal{M}$ is important because $\mathcal{M}$ can have disconnected components.

Remark 2.94. The importance of the exterior derivative in applications stems largely from the fact that it represents the coordinate-independent part of the derivative, and hence the part relevant for an objective description of the external world.

Remark 2.95. Let $\mathcal{M}$ be a manifold and (U, φ) be a chart for $\mathcal{M}$. Recall from Remark 2.41 that the local coordinates x^i are the maps $x^1 : U \to \mathbb{R} : u \to x^i(u)$ defined by the ith coefficient of $\varphi(u) = (x^1(u), \ldots, x^n(u))$ in the model space. By considering the local coordinates as functions $x^i : U \to \mathbb{R} \in \mathcal{F}(U)$ on U, the local coordinate expression for the differential

$$\mathrm{d}f = \sum_i \frac{\partial f}{\partial x^i} dx^i \tag{2.137}$$

Eq. 2.92, applied to the $x^i \in \mathcal{F}(U)$ shows that $\mathrm{d}(x^i) = dx^i$ and the dx^i are 1-forms on U, providing another definition of the dx^i. It immediately follows that $\mathrm{d}(dx^i) = \mathrm{d}(\mathrm{d}x^i) = \mathrm{d}^2 x^i = 0$ and by the Leibniz rule also $\mathrm{d}(dx^1 \wedge \ldots \wedge dx^k) = 0$.

Remark 2.96. For a finite dimensional manifold $\mathcal{M}$ with an open neighborhood U, the coordinate expression for the exterior derivative follows immediately from its definition when the coefficients are considered as functions in $\mathcal{F}(U)$, in which case it is known that the exterior derivative is given by the usual differential. By Remark 2.91, for a differential form $\alpha \in \Omega^k(U)$ we have

$$\alpha(u) = \alpha_{i_1, \cdots, i_k}(u) \, dx^{i_1} \wedge \ldots \wedge dx^{i_k} \ , \ i_1 < \ldots < i_k \tag{2.138}$$

where summation over all monotonic shuffles $i_1 < \ldots < i_k$ is implied. Using the Leibniz rule for differential forms and considering the coefficients $\alpha_{i_1, \cdots, i_k}(u)$ as functions in $\mathcal{F}(U)$ we have

$$\mathrm{d}\alpha(u) = \mathrm{d}\alpha_{i_1, \cdots, i_k}(u) \wedge (dx^{i_1} \wedge \ldots \wedge dx^{i_k})$$
$$+ (-1)^k \alpha_{i_1, \cdots, i_k}(u) \wedge \mathrm{d}(dx^{i_1} \wedge \ldots \wedge dx^{i_k}). \tag{2.139}$$

But by $\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} = 0$ and Remark 2.95 the second term vanishes. With Eq. 2.137 we hence obtain

$$\mathrm{d}\alpha(u) = \left(\frac{\partial \alpha_{i_1, \cdots, i_k}}{\partial x^j} dx^j \right)(u) \wedge (dx^{i_1} \wedge \ldots \wedge dx^{i_k}) \tag{2.140a}$$

or more explicitly

$$\mathrm{d}\alpha(u) = \sum_{i_1 < \ldots < i_k} \sum_{j \neq i_l} \left(\frac{\partial \alpha_{i_1, \cdots, i_k}}{\partial x^j} dx^j \right)(u) \wedge (dx^{i_1} \wedge \ldots \wedge dx^{i_k}) \tag{2.140b}$$

where the second sum is over all $j \neq i_l$ for $i_1 \leq \ldots \leq i_l \leq \ldots \leq i_k$.

Example 2.70. Let $\mathcal{M} = \mathbb{R}^2$ and consider the 1-form $\alpha = f(x, y) \, dx + g(x, y) \, dy$. Formally, we have $\alpha = f(x, y) \wedge dx + g(x, y) \wedge dy$. Using the linearity of the derivative and the Leibniz rule we obtain

$$\mathrm{d}\alpha = \mathrm{d}f(x, y) \wedge dx + f(x, y) \wedge \mathrm{d}(dx) + \mathrm{d}g(x, y) \wedge dy + g(x, y) \wedge \mathrm{d}(dy). \tag{2.141a}$$

But since $\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} = 0$ and by Remark 2.95 we have

$$\mathrm{d}\alpha = \mathrm{d}f(x,y) \wedge dx + \mathrm{d}g(x,y) \wedge dy. \tag{2.141b}$$

Computing the differentials of the functions, Eq. 2.137, yields

$$\mathrm{d}\alpha = \left(\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy \right) \wedge dx + \left(\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy \right) \wedge dy \tag{2.141c}$$

and since $dx \wedge dx = 0$ and $dy \wedge dy = 0$ we have

$$\mathrm{d}\alpha = \frac{\partial f}{\partial y}dy \wedge dx + \frac{\partial g}{\partial x}dx \wedge dy. \tag{2.141d}$$

With the anti-symmetry of differential forms this can be written as

$$\mathrm{d}\alpha = -\frac{\partial f}{\partial y}dx \wedge dy + \frac{\partial g}{\partial x}dx \wedge dy \tag{2.141e}$$

$$= \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx \wedge dy. \tag{2.141f}$$

Example 2.71. Let $\mathcal{M} = \mathbb{R}^3$ and the 2-form $\alpha \in \Omega^2(\mathbb{R}^3)$ be given by

$$\alpha = f(x,y,z)(dy \wedge dz) + g(x,y,z)(dz \wedge dx) + h(x,y,z)(dx \wedge dy), \tag{2.142}$$

cf. Example 2.64. Then

$$\begin{aligned} \mathrm{d}\alpha = \quad & \left(\frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz \right) \wedge (dy \wedge dz) \\ & + \left(\frac{\partial g}{\partial x}dx + \frac{\partial g}{\partial y}dy + \frac{\partial g}{\partial z}dz \right) \wedge (dz \wedge dx) \\ & + \left(\frac{\partial h}{\partial x}dx + \frac{\partial h}{\partial y}dy + \frac{\partial h}{\partial z}dz \right) \wedge (dx \wedge dy) \end{aligned} \tag{2.143a}$$

and since terms with repeated basis forms vanish one has

$$\mathrm{d}\alpha = \frac{\partial f}{\partial x}(dx \wedge dy \wedge dz) + \frac{\partial g}{\partial y}(dy \wedge dz \wedge dx) + \frac{\partial h}{\partial z}(dz \wedge dx \wedge dy). \tag{2.143b}$$

Re-ordering the basis functions while respecting anti-symmetry yields

$$\mathrm{d}\alpha = \left(\frac{\partial f}{\partial x} + \frac{\partial g}{\partial x} + \frac{\partial h}{\partial x} \right)(dx \wedge dy \wedge dz). \tag{2.143c}$$

With the Hodge dual, Def. 2.135, we obtain

$$(\star \mathrm{d}\alpha) = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial x} + \frac{\partial h}{\partial x}, \tag{2.144}$$

which is the usual expression for the divergence. Hence, using the musical isomorphisms in Def. 2.123 and Def. 2.135, we can write the divergence of a vector $u \in \mathbb{R}^3$ as

$$\operatorname{div}(u) = \nabla \cdot u = \star \mathrm{d}(\star u^\flat). \tag{2.145}$$

However, as remarked before when we considered the intrinsic definition of the cross product, when the divergence is used one is usually interested in the volume form that arises from a 2-form and not a function. The exterior calculus expression is also intrinsic and enables to extend the concept to arbitrary coordinate systems and spaces. For example, Eq. 2.141f in Example 2.70 can be seen as the divergence in $\mathbb{R}^2$, which has important applications for example in two dimensional fluid dynamics.

Example 2.72. Let $\mathcal{M} = S^2$ and consider the 1-form $\alpha = f(\theta, \phi)\, d\theta + g(\theta, \phi)\, d\phi$. Then

$$\begin{aligned}
\mathrm{d}\alpha &= \left(\frac{\partial f}{\partial \theta} d\theta + \frac{\partial f}{\partial \phi} d\phi \right) \wedge d\theta + \left(\frac{\partial g}{\partial \theta} d\theta + \frac{\partial g}{\partial \phi} d\phi \right) \wedge d\phi \\
&= \frac{\partial f}{\partial \phi} d\phi \wedge d\theta + \frac{\partial g}{\partial \theta} d\theta \wedge d\phi \\
&= \left(\frac{\partial g}{\partial \theta} - \frac{\partial f}{\partial \phi} \right) d\theta \wedge d\phi
\end{aligned}$$

and the computations in spherical coordinates are no more difficult than in Euclidean ones.

Next we will study how the exterior derivative behaves under mappings, and in particular under the pullback which is a natural operation for differential forms.[110]

Proposition 2.59. *Let* $\varphi : \mathcal{M} \to \mathcal{N}$ *be a map between manifolds* $\mathcal{M}, \mathcal{N}$. *Then the pullback* $\varphi^* : \Omega(\mathcal{N}) \to \Omega(\mathcal{M})$ *is a homomorphism of differential algebras so that*

i) $\varphi^*(\alpha \wedge \beta) = \varphi^* \alpha \wedge \varphi^* \beta$;

ii) $\varphi^*(\mathrm{d}\alpha) = \mathrm{d}(\varphi^* \alpha)$ *and the exterior derivative is natural with respect to the pullback and the following diagram commutes*

$$\begin{array}{ccc}
\Omega^k(\mathcal{M}) & \xleftarrow{\;\varphi^*\;} & \Omega^k(\mathcal{N}) \\
{\scriptstyle \mathrm{d}} \downarrow & & \downarrow {\scriptstyle \mathrm{d}} \\
\Omega^{k+1}(\mathcal{M}) & \xleftarrow[\;\varphi^*\;]{} & \Omega^{k+1}(\mathcal{N})
\end{array}$$

----

[110] For a general theory of natural operators in differential geometry see (Kolar, Michor, and Slovak, *Natural Operations in Differential Geometry*).

Remark 2.97. For the pullback it is sufficient that $\varphi : \mathcal{M} \to \mathcal{N}$ is a C^1 map and no inverse has to exist. These modest requirements are an important reason for the prevalence of pullbacks in many applications.

The exterior derivative is also natural with respect to push-forwards φ_* by diffeomorphisms and the following diagram commutes

$$
\begin{array}{ccc}
\Omega^k(\mathcal{M}) & \xrightarrow{\varphi_*} & \Omega^k(\mathcal{N}) \\
\mathrm{d}\downarrow & & \downarrow \mathrm{d} \\
\Omega^{k+1}(\mathcal{M}) & \xrightarrow[\varphi_*]{} & \Omega^{k+1}(\mathcal{N})
\end{array}
$$

The interior product, that was introduced in Def. 2.104 for tensors, place a special role in the exterior algebra where it is an anti-derivation, cf. Remark 2.55.

Definition 2.140. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$. The **interior product** $\mathrm{i}_X : \Omega^{k+1}(\mathcal{M}) \to \Omega^k(\mathcal{M})$ of $\omega \in \Omega^{k+1}(\mathcal{M})$ with the vector field X is*

$$
\mathrm{i}_X(\omega)(Y_1, \ldots, Y_k) = \omega(X, Y_1, \ldots, Y_k)
$$

for arbitrary vector fields $Y_i \in \mathfrak{X}(\mathcal{M})$, and for $\omega \in \Omega^0(\mathcal{M})$ the interior product is defined as $\mathrm{i}_X \omega = 0$.

As for tensors, the interior product is sometimes also considered as a contraction. Important properties of the interior product are summarized in the following proposition.

Proposition 2.60. *Let $\mathcal{M}$ be a manifold and $\alpha \in \Omega^k(\mathcal{M})$, $\beta \in \Omega^l(\mathcal{M})$, and $f \in \Omega^0(\mathcal{M})$. Then*

i) i_X is an anti-derivation of the exterior algebra, that is it is $\mathbb{R}$-linear and satisfies the Leibniz rule $\mathrm{i}_X(\alpha \wedge \beta) = \mathrm{i}_X\alpha \wedge \beta + (-1)^k \alpha \wedge \mathrm{i}_X\beta$;

ii) $\mathrm{i}_{fX}\alpha = f\mathrm{i}_X\alpha$;

iii) $\mathrm{i}_X\mathrm{d}f = X[f]$.

The behaviour of the interior product under mappings is summarized in the next proposition.

Proposition 2.61. *Let $\mathcal{M}$ and $\mathcal{N}$ be manifolds and $\varphi : \mathcal{M} \to \mathcal{N}$, and let $\omega \in \Omega^k(\mathcal{N})$, $X \in \mathfrak{X}(\mathcal{N})$, $Y \in \mathfrak{X}(\mathcal{M})$, and furthermore assume that X is the push-forward of Y by φ, that is $\varphi_*Y = X$. Then*

$$
\varphi^*(\mathrm{i}_X\omega) = \mathrm{i}_Y(\varphi^*\omega)
$$

and if φ is a diffeomorphism then

$$
\mathrm{i}_{\varphi_*X}(\varphi^*\omega) = \varphi^*(\mathrm{i}_X\omega).
$$

Closed and exact differential forms play an important role in the exterior algebra, and the Poincaré lemma connects these two types of forms.

Definition 2.141. *Let $\mathcal{M}$ be a manifold. A differential form $\omega \in \Omega^k(\mathcal{M})$ is **exact** if $\omega = \mathrm{d}\beta$ for some $\beta \in \Omega^{k-1}(\mathcal{M})$, and it is **closed** if $\mathrm{d}\omega = 0$.*

It immediately follows from $\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} = 0$ that every exact form is closed.

Lemma 2.1 (Poincaré lemma)**.** *Let $\mathcal{M}$ be a manifold and $\omega \in \Omega^k(\mathcal{M})$. If ω is closed then for every $m \in \mathcal{M}$ there exists a neighborhood U_m such that $\omega\mid_U \in \Omega^k(U)$ is exact.*

Some understanding of the neighborhoods U_m in the Poincaré lemma can be gained using the concept of a contractible manifold but we will not pursue these ideas here.

Remark 2.98. In the sequel we will need the concept of a differential double form, a form whose coefficient are itself differential forms.[111] Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be manifolds with local coordinates $x^1, \ldots, x^{n_1}$ and $\bar{x}^1, \ldots, \bar{x}^{n_2}$, respectively. Then a differential double form $\gamma \in \Omega^p(\mathcal{M}_1) \times \Omega^q(\mathcal{M}_2)$ of degree (p,q) on $\mathcal{M}_1 \times \mathcal{M}_2$ is in local coordinates given by

$$\gamma(x,\bar{x}) = \sum_{\substack{i_1 < \ldots < i_p \\ j_1 < \ldots < j_q}} c_{i_1,\ldots,i_p;j_1,\ldots,j_p}(dx^{i_1} \wedge \ldots \wedge dx^{i_p})(d\bar{x}^{j_1} \wedge \ldots \wedge d\bar{x}^{j_q}).$$

The double form γ can be considered as a form-valued differential form on either $\mathcal{M}_1$ or $\mathcal{M}_2$. For example, as a form on $\mathcal{M}_1$ it is

$$\gamma_x(\bar{x}) = \sum_{i_1 < \ldots < i_p} a_{i_1,\ldots,i_p}\, dx^{i_1} \wedge \ldots \wedge dx^{i_p}$$

where

$$a_i = \sum_{j_1 < \ldots < j_q} c_{i_1,\ldots,i_p;j_1,\ldots,j_p}\, dx^{j_1} \wedge \ldots \wedge dx^{j_p}.$$

Since a double form can always be regarded as an ordinary differential form by fixing one of its "legs" in either $\mathcal{M}_1$ or $\mathcal{M}_2$ all standard operation for forms carry over to double forms. One often has differential double forms defined over the same manifold and we will denote the space of such forms as $\Omega^{p,q}(\mathcal{M})$.

---

[111] Double forms were introduced into the mathematics community by de Rahm (*Differentiable Manifolds: Forms, Currents, Harmonic Forms*, p. 30) although they seem to have been in ample use in physics since at least the mid 20th century, see for example the article by Tucker Tucker, "Differential Form Valued Forms and Distributional Electromagnetic Sources" and references therein. Additional applications can for example be found in (Thirring, *A Course in Mathematical Physics 1 and 2*; Warnick and Arnold, "Electromagnetic Green Functions using Differential Forms"; Kanso et al., "On the geometric character of stress in continuum mechanics").

Volume Forms and Densities on Manifolds   Next, we will generalize the notions of volume form, orientability, and density, introduced in Def. 2.130 for linear spaces, to manifolds.

Definition 2.142. *Let $\mathcal{M}$ be a manifold with* $\dim(\mathcal{M}) = n$. *A **volume form** $\mu \in \Omega^n(\mathcal{M})$is an n-form such that $\mu(m) \neq 0$ for all $m \in \mathcal{M}$. A manifold $\mathcal{M}$ is **orientable** if there exists a volume form on $\mathcal{M}$. The tuple $(\mathcal{M}, \mu)$ is a **volume manifold**.*

An alternative characterization of an orientable manifold is the following.

Proposition 2.62. *Let $\mathcal{M}$ be a manifold with atlas $\{(U_i, \varphi_i)\}$ and $\varphi : U_i \to \bar{U}_i \subset \mathbb{R}^n$. $\mathcal{M}$ is orientable if and only if the Jacobian determinant of the overlap maps $t_{ij} = \varphi_i \circ \varphi_j^{-1} : \bar{U}_j \to \bar{U}_i$ is positive.*

As in the linear case, an orientation of a manifold is an equivalence class of volume forms, and it will be introduced next.

Definition 2.143. *Let $\mathcal{M}$ be an orientable manifold. Then two volume forms μ_1 and μ_2 are **equivalent** if there exists an $f \in \mathcal{F}(\mathcal{M})$ with $f(m) > 0$ for all $m \in \mathcal{M}$ such that $\mu_2 = f\mu_1$. An **orientation of** $\mathcal{M}$ is an equivalence class $[\mu]$ of volume forms on $\mathcal{M}$, and an **oriented manifold** is the tuple $(\mathcal{M}, [\mu])$ of a manifold $\mathcal{M}$ and an equivalence class $[\mu]$ of orientations.*

Example 2.73. The Möbius strip is non-orientable.

Proposition 2.63. *Let $\mathcal{M}$ be an orientable manifold. $\mathcal{M}$ is connected if and only if $\mathcal{M}$ has exactly two orientations.*

Volume preserving maps are central to continuum mechanics and we will formally introduce the notion next.

Definition 2.144. *Let $\mathcal{M}$ and $\mathcal{N}$ be manifolds with volume form $\mu_{\mathcal{M}}$ and $\mu_{\mathcal{N}}$. A smooth map $\varphi : \mathcal{M} \to \mathcal{N}$ is **volume preserving** if $\varphi^* \mu_{\mathcal{N}} = \mu_{\mathcal{M}}$ and it is **orientation preserving** if $\varphi^* \mu_{\mathcal{N}} = [\mu_{\mathcal{M}}]$.*

Clearly, the above definitions depend on the volume forms $\mu_{\mathcal{M}}$ and $\mu_{\mathcal{N}}$ that are chosen on the manifolds.

Proposition 2.64. *Let $\mathcal{M}, \mathcal{N}$ be manifolds with volume forms $\mu_{\mathcal{M}}$ and $\mu_{\mathcal{N}}$, respectively, and let $\varphi : \mathcal{M} \to \mathcal{N}$ be a smooth ma mapp. Then $\varphi^* \mu_{\mathcal{N}}$ is a volume form for $\mathcal{M}$ if and only if φ is a local diffeomorphism and for every $m \in \mathcal{M}$ there is a neighborhood U_m such that $\varphi \mid_{U_m} : U_m \to \varphi(U_m)$ is a diffeomorphism.*

With the above definitions we can generalize the notion of the determinant from linear spaces to manifolds.

Definition 2.145. *Let $\mathcal{M}, \mathcal{N}$ be manifolds with volume forms $\mu_{\mathcal{M}}$ and $\mu_{\mathcal{N}}$, and $\varphi : \mathcal{M} \to \mathcal{N}$ be smooth. The **Jacobian determinant** of the mapping φ with respect to $\mu_{\mathcal{M}}$ and $\mu_{\mathcal{N}}$ is the unique function $J_\varphi(\mu_{\mathcal{N}}, \mu_{\mathcal{M}})(m) \in \mathcal{F}(\mathcal{M})$ such that $\varphi^* \mu_{\mathcal{N}} = J_\varphi(\mu_{\mathcal{N}}, \mu_{\mathcal{M}}) \mu_{\mathcal{M}}$.*

Analogous to the requirement in the linear case that two vector spaces are isomorphic if and only if the determinant is non-vanishing, we have in the manifold case the following theorem, cf. also Proposition 2.64.

Proposition 2.65. *Let* $\varphi : \mathcal{M} \to \mathcal{N}$ *be a smooth map between manifolds. Then* φ *is a local diffeomorphism at* m *if and only if* $J_{\varphi}(\mu_{\mathcal{N}}, \mu_{\mathcal{M}}) \neq 0$.

Again similar to the linear case, from the definition of the pullback additional properties of the Jacobian determinant follow.

Proposition 2.66. *Let* $(\mathcal{M}, \mu)$ *be a volume manifold.*

i) If $f : \mathcal{M} \to \mathcal{M}$ *and* $g : \mathcal{M} \to \mathcal{M}$ *are smooth maps then* $J(f \circ g) = J(g)\,(J(f) \circ g)$.

ii) If $\mathrm{id} : \mathcal{M} \to \mathcal{M}$ *is the identity on* $\mathcal{M}$ *then* $J(\mathrm{id}) = 1$.

iii) If $f : \mathcal{M} \to \mathcal{M}$ *is a diffeomorphism then*

$$J(f^{-1}) = \frac{1}{J(f) \circ f^{-1}}.$$

The next proposition characterizes volume preserving mappings using the Jacobian determinant.

Proposition 2.67. *Let* $\mathcal{M}, \mathcal{N}$ *be manifolds with volume forms* $\mu_{\mathcal{M}}$ *and* $\mu_{\mathcal{N}}$, *and* $\varphi : \mathcal{M} \to \mathcal{N}$. *Then* φ *is volume preserving with respect to* $\mu_{\mathcal{M}}$ *and* $\mu_{\mathcal{N}}$ *if and only if* $J_{\varphi}(\mu_{\mathcal{M}}, \mu_{\mathcal{N}}) = 1$ *for all* $m \in \mathcal{M}$, *and it is orientation preserving if and only if* $J_{\varphi}(\mu_{\mathcal{M}}, \mu_{\mathcal{N}}) > 0$.

We will return to volume preserving mappings when we introduced the Lie derivative that will enable us to characterize them by their infinitesimal generator. Densities for vector spaces were introduced in Def. 2.131. We will now generalize the concept to manifolds.

Definition 2.146. *Let* $\mathcal{M}$ *be an* n-*dimensional manifold. The* **density bundle** $\pi_{\alpha} : |\Omega|^{\alpha}(\mathcal{M}) \to \mathcal{M}$ *is the vector bundle on* $\mathcal{M}$ *such that* $\pi_{\alpha}^{-1}(m) = |\bigwedge^{n}|^{\alpha}(T_m\mathcal{M})$ *for all* $m \in \mathcal{M}$. *An* α-**density** $\varrho^{\alpha} \in |\Omega|^{\alpha}(\mathcal{M})$ *is a section of* $\pi_{\alpha} : |\Omega|^{\alpha}(\mathcal{M}) \to \mathcal{M}$. *The space of* 1-*densities* ϱ *is denoted by* $\mathrm{Den}(\mathcal{M}) = |\Omega^{n}|^{1}(\mathcal{M})$.

It can be shown that $|\Omega|^{\alpha}(\mathcal{M})$ is a well defined, one-dimensional vector bundle over $\mathcal{M}$.

Remark 2.99. An α-density $\varrho^{\alpha} \in |\Omega|^{\alpha}(\mathcal{M})$ can be constructed from a volume element $\omega \in \Omega^{n}(\mathcal{M})$ as

$$\varrho^{\alpha}(e_1, \ldots, e_n) = |\omega\,(e_1, \ldots, e_n)|^{\alpha}.$$

Note that on an orientable manifold the space of densities $\mathrm{Den}(\mathcal{M}) = |\Omega|(\mathcal{M})$ can be canonically identified with the space of volume forms $\Omega^{n}(\mathcal{M})$.

Remark 2.100. In most situations densities behave like volume form. For example, the pullback of a density $\varrho \in \text{Den}(\mathcal{N})$ by a map $\varphi : \mathcal{M} \to \mathcal{N}$ is

$$(\varphi^* \varrho)(m) = \varrho((T_m\varphi)\, v_1, \ldots, (T_m\varphi)\, v_n)$$

for $v_i \in T_m\mathcal{M}$, cf. Def. 2.139. The critical difference between densities and volume elements becomes apparent on non-orientable manifolds. On such manifolds it is not possible to pull-back a volume form into charts such that the integrals in the overlap of the charts do agree—there will be a sign inconsistency caused by the inconsistent orientation of the charts, cf. Proposition 2.62. Since for densities the absolute value of the determinant is employed this problem is avoided and integration is well defined even when the manifold is non-orientable. The following proposition restates this in slightly different terms.

Proposition 2.68. *Let $\mathcal{M}$ be a manifold. Then there exists a smooth positive density on $\mathcal{M}$.*

Differential Forms on Riemannian Manifolds   We close this section by generalizing the concepts of a metric-induced volume forms and the Hodge dual, which were introduced in Def. 2.133 and Def. 2.135 for linear spaces, to manifolds. Recall from Def. 2.122 that a Riemannian manifold is a manifold with a non-degenerate, symmetric covariant tensor field $g \in \mathcal{T}_2^0(\mathcal{M})$ that defines a bilinear form in each tangent space.

Definition 2.147. *Let $(\mathcal{M}, g)$ be a Riemannian manifold that is orientable and for arbitrary $m \in \mathcal{M}$ let $\{\xi_1, \ldots, \xi_n\}$ be an arbitrary positively oriented, g-orthonormal basis for $T_m\mathcal{M}$ with dual basis $\{\tilde{\xi}^1, \ldots, \tilde{\xi}^n\}$.*

*i) The g-**volume** on $\mathcal{M}$ is then the unique volume element $\mu_g(m)$ such that*

$$\mu_g(\xi_1, \ldots, \xi_n) = 1,$$

and using the dual basis it is given by

$$\mu_g = \tilde{\xi}^1 \wedge \ldots \wedge \tilde{\xi}^n.$$

With local basis functions $\{\partial/\partial x^1, \ldots, \partial/\partial x^n\}$ and duals $\{dx^1, \ldots, dx^n\}$ obtained from a chart for $\mathcal{M}$ the g-volume is

$$\mu_g = |\det(g_{ij})|^{1/2}\, dx^1 \wedge \ldots \wedge dx^n.$$

*ii) The g-α-**density** $|\mu|^\alpha$ is the unique density on $\mathcal{M}$ such that*

$$|\mu|^\alpha(\xi_1, \ldots, \xi_n) = 1,$$

and using the dual basis functions it is given by

$$|\mu|^\alpha = |\tilde{\xi}^1 \wedge \ldots \wedge \tilde{\xi}^n|^\alpha.$$

The above definitions generalize in a straight-forward manner their linear analogues. The same holds for the Hodge dual. We begin by introducing the metric pairing $\langle\!\langle\,,\rangle\!\rangle$ for differential k-forms on manifolds, which is defined point-wise by the pairing for exterior k-forms in Def. 2.134.

Definition 2.148. *Let $(\mathcal{M}, g)$ be a Riemannian manifold. Then the **metric pairing***

$$\langle\!\langle\,,\rangle\!\rangle : \Omega^k(\mathcal{M}) \times \Omega^k(\mathcal{M}) \to \mathcal{F}(\mathcal{M})$$

between two differential k-forms $\alpha, \beta \in \Omega^(\mathcal{M})$ is

$$
\begin{aligned}
\langle\!\langle \alpha, \beta \rangle\!\rangle(u) &= \sum_{i_1 < \cdots < i_k} \alpha_{i_1, \cdots, i_k}(u)\, \beta^{i_1, \cdots, i_k}(u) \\
&= \sum_{i_1 < \cdots < i_k} \alpha_{i_1, \cdots i_k}(u)\, \left(g^{j_1, i_1}(u) \ldots g^{j_k, i_k}(u)\, \beta_{j_1, \cdots j_k}(u) \right)
\end{aligned}
$$

for $u \in U \subset \mathcal{M}$.

Remark 2.101. In contrast to the linear case, on manifolds some care is required when the metric pairing $\langle\!\langle\,,\rangle\!\rangle$ is interpreted as an inner product since the image space is no longer $\mathbb{R}$ but the space of functions $\mathcal{F}(\mathcal{M})$. A proper inner product can hence be defined by integrating the image of the pairing over $\mathcal{M}$, cf. Def. 2.165.

With the above pairing, the Hodge dual on manifolds can be defined as follows.

Definition 2.149. *Let $(\mathcal{M}, g)$ be an n-dimensional Riemannian manifold, and μ be the volume form induced by the metric. The **Hodge dual***

$$\star : \Omega^k(\mathcal{M}) \to \Omega^{n-k}(\mathcal{M})$$

of a differential k-form $\alpha \in \Omega^k(\mathcal{M})$ is the differential $(n-k)$-form such that $(\star\alpha)(m) = \star(\alpha(m))$ for all $m \in \mathcal{M}$ so that $\alpha \wedge \beta = \langle\!\langle \alpha, \beta \rangle\!\rangle \mu$ for $\alpha, \beta \in \Omega^k(\mathcal{M})$.

The properties from Proposition 2.58 all carry over to the manifold setting since the Hodge dual is defined fiber-wise. The co-differential is a central operator of the exterior algebra on a Riemannian manifold and it is for example employed to define the Laplace(-de Rahm) operator.

Definition 2.150. *Let (M, g) be a Riemannian manifold. The **codifferential***

$$\delta : \Omega^{k+1}(\mathcal{M}) \to \Omega^k(\mathcal{M})$$

of a $(k+1)$-form $\beta \in \Omega^{k+1}(\mathcal{M})$ is

$$\delta\beta = (-1)^{nk+1+\operatorname{ind}(g)} \star \mathrm{d} \star \beta$$

and $\delta\beta = 0$ when $\beta \in \Omega^0(\mathcal{M})$.

It follows from $\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} = 0$ that also $\delta^2 = \delta \circ \delta = 0$.

Definition 2.151. *Let (M, g) be a Riemannian manifold. A differential form $\alpha \in \Omega^k$ satisfying $\alpha = \delta\beta$ for some $\beta \in \Omega^{k+1}$ is* **co-exact**, *and a differential form $\beta \in \Omega^{k+1}(\mathcal{M})$ satisfying $\delta\beta = 0$ is* **co-closed**.

Example 2.74. Let $\mathcal{M} = \mathbb{R}^3$ and consider the differential 2-form

$$\alpha = \alpha_1(dy \wedge dz) + \alpha_2(dz \wedge dx) + \alpha_3(dx \wedge dy) \in \Omega^2(\mathbb{R}^3). \qquad (2.146)$$

Resolving the codifferential $\delta = \star \mathrm{d} \star$ operator by operator, we have for the first Hodge dual

$$\star\alpha = \alpha_1 dx + \alpha_2 dy + \alpha_3 dz. \qquad (2.147)$$

Together with the exterior derivative we thus obtain

$$\mathrm{d}(\star\alpha) = \left(\frac{\partial \alpha_3}{\partial y} - \frac{\partial \alpha_2}{\partial z} \right) dy \wedge dz \qquad (2.148)$$

$$+ \left(\frac{\partial \alpha_3}{\partial x} - \frac{\partial \alpha_1}{\partial z} \right) dx \wedge dz + \left(\frac{\partial \alpha_2}{\partial x} - \frac{\partial \alpha_1}{\partial y} \right) dx \wedge dy. \quad (2.149)$$

Finally, taking the second Hodge dual into account, and since

$$(-1)^{nk+1+\mathrm{ind}\,(g)} = (-1)^{3\cdot 1+1+0} = (-1)^4 = 1 \qquad (2.150)$$

we have

$$\delta\alpha = \star\mathrm{d}(\star\alpha) = \left(\frac{\partial \alpha_3}{\partial y} - \frac{\partial \alpha_2}{\partial z} \right) dx - \left(\frac{\partial \alpha_3}{\partial x} - \frac{\partial \alpha_1}{\partial z} \right) dy + \left(\frac{\partial \alpha_2}{\partial x} - \frac{\partial \alpha_1}{\partial y} \right) dz$$
$$(2.151)$$

which is indeed a 1-form.

Remark 2.102. Using the codifferential, alternative expressions for the standard operators from vector calculus can be obtained that do not (explicitly) depend on the exterior derivative:

$$\nabla f = \mathrm{grad}(f) = (\star\delta \star f)^\sharp \qquad (2.152\text{a})$$

$$\nabla \times X = \mathrm{curl}(X) = (\delta \star X^\flat)^\sharp \qquad (2.152\text{b})$$

$$\nabla \cdot X = \mathrm{div}(X) = -\delta X^\flat \qquad (2.152\text{c})$$

where $X \in \mathfrak{X}(\mathbb{R}^3)$ and $f \in \mathcal{F}(\mathbb{R}^3)$.

The codifferential allows us to introduce the Laplace-de Rahm and Laplace-Beltrami operators that are generalizations of the classical Laplacian.

Definition 2.152. *The **Laplace-de Rahm operator** $\Delta : \Omega^k(\mathcal{M}) \to \Omega^k(\mathcal{M})$ on a Riemannian manifold $(\mathcal{M}.g)$ is*

$$\Delta = \mathrm{d}\delta + \delta\mathrm{d}.$$

*For functions $f \in \Omega^0(\mathcal{M})$ the Laplace-de Rahm operator is known as **Laplace-Beltrami operator** and it is given by $\Delta f = \mathrm{d}\delta f + \delta\mathrm{d}f = \delta\mathrm{d}f$.*

Example 2.75. In $\mathbb{R}^3$ one has for the Laplace-Beltrami operator $\Delta f = \star\mathrm{d}\star\mathrm{d}f$ which, with the expressions for the divergence and the gradient in Eq. 2.145 and Example 2.59 or from Remark 2.102, is $\Delta f = \mathrm{div}\,\mathrm{grad}\,f$, the usual expression for the Laplacian.

Definition 2.153. *A differential form $\alpha \in \Omega^k(\mathcal{M})$ on a Riemannian manifold $(\mathcal{M}, g)$ is **harmonic** if $\Delta\alpha = 0$ and it is in the kernel of the Laplace-Beltrami operator Δ. The space of all harmonic forms of degree k is*

$$\mathcal{H}^k(\mathcal{M}) = \left\{ \alpha \in \Omega^k(\mathcal{M}) \mid \Delta\alpha = 0 \right\}.$$

2.3.2.5 The Lie Derivative

In this section we will introduce the Lie derivative that will generalize the directional derivative of functions in Def. 2.96 to arbitrary tensor fields. In continuum mechanics, the Lie derivative is central to the description of transport processes, and hence flows of vector fields will also again be of relevance.

The Lie Derivative for Functions and Vector Fields   We will begin by studying the transport of functions along the flow of a vector field on a manifold. A first question of considerable interest is which function spaces are closed under such a flow. For $\mathbb{R}^n$ and diffeomorphisms $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ that are the identity sufficiently large away from the origin an answer is provided by the following theorem.

Theorem 2.17 (Triebel).[112] *Let $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism that satisfied $\varphi(x) = x$ for large $|x|$ and that acts on functions by the pullback $\varphi^* f = f \circ \varphi$, and let $B^s_{p,q}(\mathbb{R}^n)$ be the Besov space on $\mathbb{R}^n$ for $0 < p \le \infty$, $0 < q \le \infty$, and $-\infty < s < \infty$, and $F^s_{q,p}(\mathbb{R}^n)$ be the Triebel-Lizorkin space on $\mathbb{R}^n$ for $0 < p < \infty$, $0 < q \le \infty$, and $-\infty < s < \infty$. Then φ is a one-to-one mapping from $B^s_{q,p}(\mathbb{R}^n)$ onto itself, and a one-to-one mapping from $F^s_{q,p}(R^n)$ onto itself.*

The Besov spaces $B^s_{q,p}$ and Triebel-Lizorkin spaces $F^s_{q,p}$ subsume all standard function spaces such as Lebesgue, Sobolev, Hardy, and for example $B^s_{2,2}(\mathbb{R}^n)$ are the classical Sobolev-Hilbert spaces $H^s(\mathbb{R}^n)$. The above result has hence great generality on $\mathbb{R}^n$. Unfortunately, to our knowledge no results for the manifold case exist.

For manifolds, it is known that the pullback of smooth functions $\mathcal{F}(\mathcal{M})$ is a function algebra automorphism of $\mathcal{F}(\mathcal{M})$, that is $F_t^*(f\,g) = (F_t^* f)\,(F_t^* g)$ for $f, g \in \mathcal{F}(\mathcal{M})$. The following theorem shows that also the converse holds.

---

[112]Triebel, *Theory of Function Spaces*, Chapter 2.10.2.

Theorem 2.18. *Let $\mathcal{M}$ be a compact, finite dimensional manifold. Then every invertible, linear mapping $\psi : \mathcal{F}(\mathcal{M}) \to \mathcal{F}(\mathcal{M})$ that is an automorphism of $\mathcal{F}(\mathcal{M})$ satisfying $\psi(f\,g) = \psi(f)\,\psi(g)$ for $f, g \in \mathcal{F}(\mathcal{M})$ corresponds to a unique diffeomorphism $\varphi : \mathcal{M} \to \mathcal{M}$ such that $\psi(f) = f \circ \varphi$.*

Hence every automorphism of the algebra $\mathcal{F}(\mathcal{M})$ is generated by a diffeomorphism. We are not aware of an analogous result for Besov or Triebel-Lizorkin spaces.

Globally, the transport of a function by a vector field $X \in \mathfrak{X}(\mathcal{M})$ with flow F_t on a manifold $\mathcal{M}$ is described by the pullback $F_t^* f = f \circ F_t$, and when X is global vector field then F_t is a diffeomorphism by Corollary 2.9. Locally, the transport is described by the directional derivative $X[f]$ which hence provides the Lie derivative for functions.

Definition 2.154. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$ be a vector field on $\mathcal{M}$. The **Lie derivative** $\pounds_X : \mathcal{F}(\mathcal{M}) \to \mathcal{F}(\mathcal{M})$ **of a function** $f \in \mathcal{F}(\mathcal{M})$ along X is*

$$\pounds_X f = X[f] = \mathrm{d}f(X).$$

For a function, the Lie derivative hence coincides with the directional derivative introduced in Def. 2.96. The connection to the pullback is formalized in the following proposition.

Proposition 2.69. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$ be a vector field on $\mathcal{M}$ with flow F_t. Then for a function $f \in \mathcal{F}(\mathcal{M})$ it holds*

$$\frac{d}{dt} F_t^* f = F_t^*(\pounds_X f) = F_t^*(X[f]),$$

or, equivalently,

$$\left.\frac{d}{dt}\right|_{t=0} F_t^* f = \pounds_X f.$$

It is important to note that the pullback $F_t^* f = (F_t^* f)(m) = f(F_t m)$ is always defined at the same point $m \in \mathcal{M}$ and it is a time varying function at m. The time derivative $d/dt(F_t^* f)$ is thus the usual derivative of a one-dimensional function.

Remark 2.103. Let $\mathcal{M} = \mathbb{R}^3$, and $X \in \mathfrak{X}(\mathbb{R}^3)$ a time-invariant vector field with flow $F_t = x(t) : [a, b] \to \mathbb{R}^3$. Using the chain rule and the definition of the pullback for functions, we obtain with Proposition 2.69 that

$$\pounds_X f = \frac{d}{dt} F_t^* f = \frac{d}{dt}(f \circ F_t) = \frac{\partial f}{\partial x} \frac{\partial x}{\partial t}$$

which is the usual expression for the directional derivative of f along the velocity vector field $u = \partial x / \partial t$, often written as $u \cdot \nabla f$, and coincides with the definition in Def. 2.154
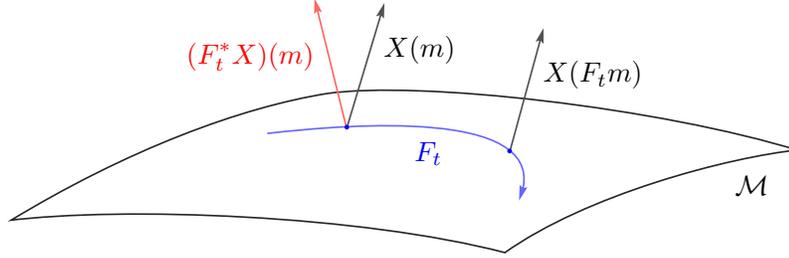
Figure 2.36: The Lie derivative $\mathcal{L}_Y X$ of a vector field $X \in \mathfrak{X}(M)$ is defined as the infinitesimal change $d/dt(F_t^* X)_{t=0}$ of X over time along the flow F_t generated by the vector field $Y \in \mathfrak{X}(\mathcal{M})$.

The Lie derivative for vector fields is defined using the Jacobi-Lie bracket $[X, Y] = X[Y[\cdot]] - Y[X[\cdot]]$ that was introduced in Def. 2.97.

Definition 2.155. *Let $\mathcal{M}$ be a manifold, and let $X, Y \in \mathfrak{X}(\mathcal{M})$ be vector fields on $\mathcal{M}$ with F_t being the flow of X. The **Lie derivative** $\mathcal{L}_X : \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M})$ **of** Y **along** X is*

$$\mathcal{L}_X Y = [X, Y] = -[Y, X].$$

The connection between the Lie derivative and the pullback of vector fields is established in the following proposition.

Proposition 2.70. *Let $\mathcal{M}$ be a manifold and $X, Y \in \mathfrak{X}(\mathcal{M})$ vector fields on $\mathcal{M}$, and let F_t be the flow of X. Then*

$$\frac{d}{dt} F_t^* Y = F_t^*(\mathcal{L}_X Y),$$

or, equivalently,

$$\left. \frac{d}{dt} \right|_{t=0} F_t^* Y = \mathcal{L}_X Y.$$

Analogous to the situation for functions, $F_t^* Y$ is a time varying vector field in the fiber over $m \in \mathcal{M}$ for all t, see Fig. 2.36.

Definition 2.156. *Let $\mathcal{M}$ be a manifold and $X, Y \in \mathfrak{X}(\mathcal{M})$ vector fields on $\mathcal{M}$. Then X **and** Y **commute** when $[X, Y] = 0$.*

The following proposition characterizes commuting vector fields further.

Proposition 2.71. *Let $\mathcal{M}$ be a manifold and $X, Y \in \mathfrak{X}(\mathcal{M})$ vector fields on $\mathcal{M}$ and F_t and G_t their respective flows. Then the following are equivalent:*

i) $[X, Y] = 0$;

ii) $F_t^* Y = Y$;

iii) $G_t^* X = X$;

iv) $F_t \circ G_t = G_t \circ F_t$.

The above results for the Lie derivative of static vector fields generalize in a straight forward manner to the time-dependent case.

The Lie Derivative for Tensor Fields The Lie derivative for arbitrary tensor fields is defined analogous to the Lie derivative for functions and vector fields.

Definition 2.157. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$ be a vector field on $\mathcal{M}$ with flow F_t. The **Lie derivative** $\pounds_X : \mathcal{T}_s^r(\mathcal{M}) \to \mathcal{T}_s^r(\mathcal{M})$ **of a tensor field** $\tau \in \mathcal{T}_s^r(\mathcal{M})$ **along** X is*

$$\pounds_X \tau = \left.\frac{d}{dt}\right|_{t=0} F_t^* \tau,$$

or, equivalently,

$$F_t^* \pounds_X \tau = \frac{d}{dt} F_t^* \tau.$$

For a time varying tensor field $\tau_t \in \mathcal{T}_s^r(\mathcal{M}) \times \mathbb{R}$ the Lie derivative is defined as

$$\left.\frac{d}{dt}\right|_{t=0} F_t^* \tau_t = \frac{\partial \tau_t}{\partial t} + \pounds_X \tau_t$$

or, equivalently,

$$\frac{d}{dt} F_t^* \tau_t = F_t^* \left(\frac{\partial \tau_t}{\partial t} + \pounds_X \tau_t \right).$$

The pullback $F_t^* \tau$ is again a time-varying tensor in the fiber at $m \in \mathcal{M}$ for all t, see Fig. 2.37.

Remark 2.104. In the foregoing we employed the flow of a vector field to define the Lie derivative, an ansatz which is known in the literature as dynamic definition. An alternative approach is purely algebraic and one requires that the Lie derivative is a derivation on the tensor algebra $\mathcal{T}(\mathcal{M})$, $\mathbb{R}$-linear and satisfying a generalized $r + s$ term Leibniz rule, and natural with respect to restrictions so that for $U \subset V \subset \mathcal{M}$ the following diagram commutes

$$
\begin{array}{ccc}
\mathcal{T}_s^r(V) & \xrightarrow{\ |_U\ } & \mathcal{T}_s^r(U) \\
\pounds_X \downarrow & & \downarrow \pounds_X \\
\mathcal{T}_s^r(V) & \xrightarrow[\ |_U\ ]{} & \mathcal{T}_s^r(U)
\end{array}
\qquad (2.153)
$$
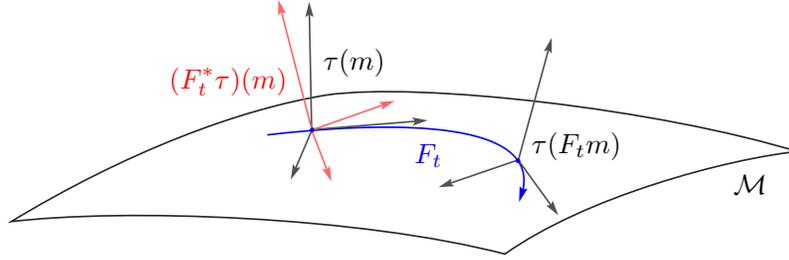
Figure 2.37: The Lie derivative $\pounds_Y\tau$ of a tensor $\tau \in \mathcal{T}_s^r(\mathcal{M})$ is defined as the infinitesimal change $d/dt(F_t^*\tau)_{t=0}$ of τ over time along the flow F_t generated by the vector field $Y \in \mathfrak{X}(\mathcal{M})$.

For example, with this approach one derives the Lie derivative for a covector field as follows. Let a function $f = (\alpha \cdot Y) \in \mathcal{F}(\mathcal{M})$ be given by the pairing of a vector field $Y \in \mathfrak{X}(\mathcal{M})$ and a covector field $\alpha \in \mathfrak{X}^*(\mathcal{M})$. Since the Lie derivative of a function is the directional derivative we obtain using the required Leibniz rule that

$$\pounds_X(\alpha \cdot Y) = \pounds_X(\alpha) \cdot Y + \alpha \cdot \pounds_X(Y).$$

Hence, the Lie derivative of the covector field α along X is

$$\pounds_X(\alpha) \cdot Y = \pounds_X(\alpha \cdot Y) - \alpha \cdot \pounds_X(Y),$$

which is well defined since it only requires the Lie derivatives for functions and vector fields. The approach can be extended to arbitrary (r, s) tensor fields. That the algebraic approach to the Lie derivative is well defined is guaranteed by the following proposition.

Proposition 2.72. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$. Then the Lie derivative $\pounds_X : \mathcal{T}_s^r(\mathcal{M}) \to \mathcal{T}_s^r(\mathcal{M})$ is the unique differential operator on $\mathcal{T}(\mathcal{M})$ such that it agrees with the definition of the Lie derivative for functions in $\mathcal{F}(\mathcal{M}) \cong \mathcal{T}_0^0(\mathcal{M})$ and vector fields in $\mathfrak{X}(\mathcal{M}) \cong \mathcal{T}_0^1(\mathcal{M})$.*

The Lie derivative is natural with respect to the push-forward, and this distinguishes it from derivatives such as the covariant derivative that can be defined on the tensor algebra.

Proposition 2.73. *Let $\varphi : \mathcal{M} \to \mathcal{N}$ be a diffeomorphism of manifolds and $X \in \mathfrak{X}(\mathcal{M})$ be a vector field on $\mathcal{M}$. The Lie derivative $\pounds_X$ is natural with respect to the push-forward so that for any tensor field $\tau \in \mathcal{T}_s^r(\mathcal{M})$ it holds*

$$\pounds_{\varphi_* X} \varphi_* \tau = \varphi_*(\pounds_X \tau).$$

Since the pullback φ^* of a tensor field is defined as $\varphi^* = (\varphi^{-1})_*$, the Lie derivative is also natural with respect to the pullback.

Corollary 2.11. *Let $\varphi : \mathcal{M} \to \mathcal{N}$ be a smooth map between manifolds and $X \in \mathfrak{X}(\mathcal{M})$ and $Y \in \mathfrak{X}(\mathcal{N})$ be φ-related $X \sim_\varphi Y$. Then for a covariant tensor field $\tau \in \mathcal{T}_s^0(\mathcal{M})$ one has*

$$\varphi^*(\pounds_Y \tau) = \pounds_X \varphi^* \tau.$$

Remark 2.105. Let $\mathcal{M}$ be a finite dimensional manifold, $X, Y \in \mathfrak{X}(\mathcal{M})$ be vector fields, and $f \in \mathcal{F}(\mathcal{M})$ be a function on $\mathcal{M}$. The coordinate expressions for the Lie derivative of functions and vector fields are immediately given by those for the directional derivative and the Jacobi-Lie bracket, cf. Remark 2.54 and Remark 2.57. Hence,

$$\pounds_X f = X^i \frac{\partial f}{\partial x^i}$$

and

$$(\pounds_X Y)^i = [X, Y]^i = X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}$$

where, as usual, we identified objects and their local representatives. The coordinate expression for the Lie derivative of a covector $\alpha \in \mathfrak{X}(\mathcal{M})$ can be obtained by combining the definitions of the Lie derivative and of the pullback,

$$\pounds_X \alpha = \left. \frac{d}{dt} \right|_{t=0} F_t^* \alpha$$

$$= \left. \frac{d}{dt} \right|_{t=0} \alpha_i(F_t) \frac{\partial F_t^i}{\partial x^j} dx^j$$

where F_t^i are the components of F_t and $\partial F_t^i / \partial x^j$ is the tangent of the flow map. Using the Leibniz rule for the derivative with respect to time for each term in the implicit sum yields

$$\pounds_X \alpha = \left(\frac{\partial \alpha_i}{\partial x^k} \frac{\partial F_t^k}{\partial t} \frac{\partial F_t^i}{\partial x^j} dx^j + \alpha_i \frac{\partial X^i}{\partial x^j} dx^j \right)_{t=0}$$

where in the last term we also used that partial derivatives commute and that $\partial \varphi / \partial t = X$. $F_t|_{t=0}$ is the identity and hence $\partial F_t^i / \partial x^j = \delta_j^i$ so that

$$\pounds_X \alpha = \frac{\partial \alpha_i}{\partial x^k} X^k \delta_j^i dx^j + \alpha_i \frac{\partial X^i}{\partial x^j} dx^j$$

which yields for the components of the Lie derivative of a covector

$$(\pounds_X \alpha)_j = \left(\frac{\partial \alpha_j}{\partial x^k} X^k + \alpha_i \frac{\partial X^i}{\partial x^j} \right) dx^j.$$

The same ansatz can be used to derive coordinate expressions for arbitrary (r, s) tensors. Alternatively, since (r, s) tensors act on r covectors and s vectors one can employ the derivation property of the Lie derivative, which takes the form of an iterated Leibniz rule, and the above expressions for Lie derivative of vectors and covectors.

The Lie Derivative of Differential Forms We will now restrict our attention to the Lie derivative of differential forms. Similar to other constructions, the additional structure provided by the exterior complex will make the Lie derivative an even more powerful tool.

Definition 2.158. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$ a vector field on $\mathcal{M}$ with flow F_t. The **Lie derivative** $\pounds_X : \Omega^k(\mathcal{M}) \to \Omega^k(\mathcal{M})$ **of a differential form** $\alpha \in \Omega^k(\mathcal{M})$ **along X** is*

$$\pounds_X \alpha = \left. \frac{d}{dt} \right|_{t=0} F_t^* \alpha$$

or, equivalently,

$$F_t^* \pounds_X \alpha = \frac{d}{dt} F_t^* \alpha.$$

The Lie derivative is natural with respect to exterior differentiation, which follows immediately from Proposition 2.59, ii)

Corollary 2.12. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$ a vector field on $\mathcal{M}$, and let $\alpha \in \Omega^k(\mathcal{M})$ be a differential k-form. Then*

$$\mathrm{d}\pounds_X \omega = \pounds_X \mathrm{d}\omega.$$

Additional properties of the Lie derivative of differential forms are summarized in the following proposition.

Proposition 2.74. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$ be a vector field on $\mathcal{M}$. Then for any $\alpha \in \Omega^k(\mathcal{M})$, $\beta \in \Omega^l(\mathcal{M})$, $f \in \mathcal{F}(\mathcal{M})$ one has*

i) $\pounds_X(\alpha \wedge \beta) = \pounds_X \alpha \wedge \beta + \alpha \wedge \pounds_X \beta$;

*ii) **Cartan's formula**: $\pounds_X \alpha = \mathrm{i}_X \mathrm{d}\alpha + \mathrm{d}\mathrm{i}_X \alpha$;*

iii) $\pounds_{fX} \alpha = f \pounds_X \alpha$.

Remark 2.106. Cartan's formula is particularly useful and important in applications. For example, since $\mathrm{i}_X f = 0$ it implies that the Lie derivative of a function is

$$\pounds_X f = \mathrm{i}_X \mathrm{d}f \tag{2.154}$$

which is, as expect, the expression for the directional derivative of a function along X. Conversely, since $\mathrm{d}\mu = 0$ for a volume form μ, the Lie derivative becomes in this case

$$\pounds_X \mu = \mathrm{d}\mathrm{i}_X \mu. \tag{2.155}$$

As will be discussed in more detail in the sequel, conserved quantities play a central role in physics. The concept of invariance along the flow of a vector field will therefore be of considerable importance in the following.

Definition 2.159. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$ a vector field on $\mathcal{M}$. A **differential form** $\alpha \in \Omega^k(\mathcal{M})$ **is invariant under the flow of** X when*

$$\pounds_X \alpha = 0.$$

Useful facts about invariant forms are summarized in the following theorem.

Theorem 2.19. *Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(M)$ be a vector field on $\mathcal{M}$, and let $\alpha \in \Omega^k(\mathcal{M})$ and $\beta \in \Omega^l(\mathcal{M})$ be invariant forms. Then*

i) $i_X \alpha$ *is an invariant form under X;*

ii) $\mathrm{d}\alpha$ *is an invariant form under X;*

iii) $\pounds_X \alpha$ *is closed if and only if $\mathrm{d}\alpha$ is an invariant form;*

iv) $\alpha \wedge \beta$ *is an invariant form under X.*

Additionally, it can be shown that the invariant forms form a $\wedge$-subalgebra of $\Omega(\mathcal{M})$ that is closed under d and i_X.

Divergence The special form of the Lie derivative for volume forms in Eq. 2.155 suggests the following generalized definition of divergence that does only require a volume form, in contrast to those in Eq. 2.145 requiring a metric.

Definition 2.160. *Let $(\mathcal{M}, \mu)$ be a volume manifold and $X \in \mathfrak{X}(\mathcal{M})$ be a vector field on $\mathcal{M}$. The **divergence of** X **with respect to** μ is the unique function $\mathrm{div}_\mu X \in \mathcal{F}(\mathcal{M})$ defined by*

$$\pounds_X \mu = (\mathrm{div}_\mu X)\mu.$$

*A vector field X is **incompressible** or **solenoidal** or **divergence free** if $\mathrm{div}_\mu X = 0$.*

Remark 2.107. In $\mathbb{R}^3$ with the standard volume form $dx^1 \wedge dx^2 \wedge dx^3$ and using that $\pounds_X \mu = (\mathrm{div}_\mu X)\mu = \mathrm{d}i_X \mu$ one has for a vector field

$$X = X^1 \frac{\partial}{\partial x^1} + X^2 \frac{\partial}{\partial x^2} + X^3 \frac{\partial}{\partial x^3} \qquad (2.156)$$

that

$$\mathrm{i}_X \mu = X^1(dx^2 \wedge dx^3) - X^2(dx^1 \wedge dx^3) + X^3(dx^1 \wedge dx^3) \qquad (2.157)$$

where we used repeatedly the Leibniz rule for the interior product in Proposition 2.60 and that $dx^i(\partial/\partial x^j) = \delta^i_j$. With the expression for the exterior derivative of a 2-form in $\mathbb{R}^3$ that we already obtained in Example 2.71 one has

$$\mathrm{d}i_X \mu = \left(\frac{\partial X^1}{\partial x^1} + \frac{\partial X^2}{\partial x^2} + \frac{\partial X^3}{\partial x^3} \right)(dx^1 \wedge dx^2 \wedge dx^3) \qquad (2.158)$$

so that the unique divergence function is given by

$$\mathrm{div}_\mu X = \frac{\partial X^1}{\partial x^1} + \frac{\partial X^2}{\partial x^2} + \frac{\partial X^3}{\partial x^3} \qquad (2.159)$$

which is the usual expression for the divergence of a vector field.

The following proposition characterizes the relationship between incompressibility of a vector field and properties of the associated flow.

Proposition 2.75. *Let $(\mathcal{M}, \mu)$ be a volume manifold, and $X \in \mathfrak{X}(\mathcal{M})$ be a vector field on $\mathcal{M}$ with flow F_t. Then the following are equivalent:*

i) $\mathrm{div}_\mu(X) = 0$;

ii) $J_\mu(F_t) = 1$, for all t;

iii) $F_t^* \mu = \mu$, for all t.

If X is not complete then the above equivalences hold on the domain of X.

Many volume forms in applications are of the form $f\,\mu$ where $f \in \mathcal{F}(\mathcal{M})$ is a function and $\mu \in \Omega^n(\mathcal{M})$ is a "canonical" volume form on $\mathcal{M}$. The divergence and transport of such volume forms is characterized in the following proposition.

Proposition 2.76. *Let $(\mathcal{M}, \mu)$ be a volume manifold and X a vector field on $\mathcal{M}$. Then for any non-vanishing $f \in \mathcal{F}(\mathcal{M})$*

$$\mathrm{div}_{f\mu} X = \mathrm{div}_\mu X + \frac{1}{f}\pounds_X f$$

and for $g \in \mathcal{F}(\mathcal{M})$ one has

$$\mathrm{div}_\mu gX = g\,\mathrm{div}_\mu X + \pounds_X g.$$

Remark 2.108. An definition of divergence analogous to those in Def. 2.160 for 1-densities is possible. For example, the Lie derivative of $\varrho \in \mathrm{Den}(\mathcal{M})$ on a manifold $\mathcal{M}$ along a vector field $X \in \mathfrak{X}(\mathcal{M})$ with flow F_t is

$$\pounds_X \varrho = \left.\frac{d}{dt}\right|_{t=0} F_t^* \varrho,$$

cf. Remark 2.100.

2.3.2.6 Integration on Manifolds

In the following, we will discuss how differential forms can be integrated and substantiate our previous claim that all integrands are differential forms. Our ansatz will be similar to how we previously introduced other concepts: we will pull-back a differential k-form from a k-manifold into a suitable chart where the usual notions of integration, such as the Riemann or Lebesgue integral,

depending on one's taste and the need for mathematical sophistication, can be employed. In this section we will also introduce Stokes' theorem that is central to many applications in physics and engineering, and which beautifully unifies the fundamental theorem of calculus and the theorems of Green, Stokes, and Gauss from vector calculus.

Remark 2.109. Integration over infinite dimensional manifolds requires functional integration, an area with many open problems. We will therefore restrict ourselves to integration over finite dimensional manifolds.[113]

Integrals of Differential Forms Integration of differential k-forms always requires a k-dimensional manifold as domain, and the problem can then be reduced to a linear one using the charts of the manifold.

Remark 2.110. We will recall some notions of integration over $\mathbb{R}^n$ with standard basis $\{e_1, \dots e_n\}$ and dual basis $\{e^1, \dots e^n\}$, so that the induced g-volume form is $\mathrm{vol}(R^n) = e^1 \wedge \dots \wedge e^n$. Let $f(x) \in \mathcal{F}(\mathbb{R}^n)$ be a function with compact support over an open subset $U \subset \mathbb{R}^n$. The volume form $\omega \in \Omega^n(\mathbb{R}^n)$ naturally associated with $f(x)$ is

$$\omega = f(x) \wedge e^1 \wedge \dots \wedge e^n = f(x) \wedge dx^1 \wedge \dots \wedge dx^n = f(x)\,dx^1 \dots dx^n \quad (2.160)$$

where the last notation is justified by the tensor product structure and since we consider the canonical g-volume on $\mathbb{R}^n$. The function $f(x)$ then formally corresponds to $f(x) \equiv \omega_{1,\cdots,n}(x) = \omega(x)(e_1, \dots, e_n)$. The integral of ω over U is hence

$$\int_U \omega = \int_{\mathbb{R}^n} \omega_{1,\cdots n}(x)\, e^1 \wedge \dots e^n = \int_{\mathbb{R}^n} f(x)\,dx^1 \dots dx^n \quad (2.161)$$

where the right hand side is a known integral that can be evaluated for example using Riemann or Lebesgue integration.

The change of variables formula in $\mathbb{R}^n$ states that if $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is a vector space isomorphism then

$$\int_{\mathbb{R}^n} f(x)\,dx^1 \dots dx^n = \int_{\mathbb{R}^n} |J_\varphi(x)|(f \circ \varphi^{-1})(x)\,dx^1 \dots dx^n \quad (2.162)$$

where J_φ is the Jacobian determinant of the diffeomorphism φ with respect to the standard volume form on $\mathbb{R}^n$. Using the pullback, the change of variables formula can be written for exterior forms defined over $\mathbb{R}^n$ as follows.

Theorem 2.20 (Changes of Variables for $\mathbb{R}^n$)**.** *Let* $\omega \in \bigwedge^k(V)$ *and* $\varphi : U \to V$ *be an orientation preserving diffeomorphism of open subsets* $U, V \subset \mathbb{R}^n$ *satisfying* $\varphi(U) \subset V$*. Then*

$$\int_V \omega = \int_U \varphi^*\omega.$$

---

[113]For an introduction to functional integration see for example (Cartier and DeWitt-Morette, *Functional Integration*). It should be noted that while mathematicians are still troubled by the concept, physicists employ them frequently in practice.

We will begin the definition of the integral for differential forms by defining the integral over a chart neighborhood.

Definition 2.161. *Let $\mathcal{M}$ be an orientable n-manifold and (U, φ) be a chart of $\mathcal{M}$ with the chart map $\varphi : U \subset \mathcal{M} \to \varphi(U) \subset \mathbb{R}^n$ being orientation preserving. The **integral of a compactly supported differential form** $\omega \in \Omega^n(\mathcal{M})$ **over** U with $\mathrm{supp}\,(\omega) \subset U$ is*

$$\int_{\mathcal{M}} \omega = \int_U \omega = \int_{\varphi(U)} \varphi_*(\omega|_U) = \int_{\varphi(U)} \left(\varphi^{-1}\right)^* (\omega|_U).$$

The right hand side of the above equation is an integral over $\mathbb{R}^n$,

$$\int_{\varphi(U)} \left(\varphi^{-1}\right)^* (\omega|_U) = \int_{\mathbb{R}^n} ((\left(\varphi^{-1}\right)^* (\omega|_U))_{1,\ldots,n}\, dx^1 \ldots dx^n,$$

that can be evaluated using standard integration theory for $\mathbb{R}^n$, as discussed previously. The push-forward and the pullback in the above definition agree since the inverse φ^{-1} of φ exists so that the inverse function theorem asserts that φ is a local diffeomorphism. It can be shown that the so defined integral over a chart neighborhood of a manifold is independent of the chart. The local definition above is extended to all of $\mathcal{M}$ using a partition of unity as introduced in Def. 2.120. For the integral of a differential form over a manifold we hence obtain.

Definition 2.162. *Let $\mathcal{M}$ be an n-manifold, $\mathcal{A}$ an atlas for $\mathcal{M}$, and $\{V_i, g_i\}$ be an orientation-preserving partition of unity subordinate to $\mathcal{A}$. For a differential n-form $\omega \in \Omega^n(\mathcal{M})$ the partition of unity induces a set of compactly supported n-forms $\omega_i = g_i\,\omega$. The **integral of** ω **over** $\mathcal{M}$ is then*

$$\int_{\mathcal{M}} \omega = \sum_i \int_{V_i} \omega_i.$$

Each term in the sum on the right hand side of the above formula is defined by Def. 2.161 and the sum contains only a finite number of terms since the covering of $\mathcal{M}$ induced by the partition of unity is locally finite. It can be shown that the definition is independent of the atlas and the partition of unity, and hence well defined.

With the notion of integration for differential forms, we can generalize the change of variables theorem to manifolds.

Theorem 2.21 (Change of Variables Theorem)**.** *Let $\mathcal{M}$ and $\mathcal{N}$ be oriented n-manifolds and $\varphi : \mathcal{M} \to \mathcal{N}$ an orientation-preserving diffeomorphism. For a compactly supported differential form $\omega \in \Omega^n(\mathcal{N})$ the **change of variables** formula is*

$$\int_{\mathcal{N}} \omega = \int_{\mathcal{M}} \varphi^*\omega.$$

Remark 2.111. With slight abuse of notation, one often writes for the change of variables formula

$$\int_{\varphi(\mathcal{M})} \omega = \int_{\mathcal{M}} \varphi^* \omega. \tag{2.163}$$

The definition of the integral of a function over a manifold is analogous to Eq. 2.160

Definition 2.163. *Let* $(\mathcal{M}, \mu)$ *be a volume manifold. The **integral of the function** $f \in \mathcal{F}(\mathcal{M})$ **with respect to** μ is $\int_{\mathcal{M}} f\mu$.*

The compatibility of the above definition of the integral of a function with classical measures theory is shown in the following theorem.

Theorem 2.22. *Let* $(\mathcal{M}, \mu)$ *be a volume manifold and Σ the collection of Borel sets of $\mathcal{M}$, that is the σ-algebra generated by the open subsets of $\mathcal{M}$. Then there is a unique measure m_μ such that for every continuous function of compact support*

$$\int_{\mathcal{M}} f \, m_\mu = \int_{\mathcal{M}} f\mu.$$

Remark 2.112. An integral for any manifold, orientable or not, can be defined analogously to the above definitions using 1-densities, cf. Def. 2.146 and the ensuing discussion, when Jacobians are replaced by their absolute value. See also Remark 2.108 on the definition of the Lie derivative for densities.

In many applications one has a differential k-form defined over an n-manifold, with $k < n$, and one is interested in the integral over a k-dimensional submanifold. Exploiting that the pullback does not require an invertible map, we can thereby employ a map from a neighborhood of Euclidean space onto the (sub-)manifold to obtain a more "robust" notion of integration.

Definition 2.164. *Let $\mathcal{M}$ be an orientable n-manifold and $\psi : V \subset \mathbb{R}^k \to U \subset \mathcal{M}$ be a differentiable (not necessarily full rank) map between open sets $V \subset \mathbb{R}^k$ and $U \subset \mathcal{M}$ that is orientation preserving. The integral of a k-form with* $\operatorname{supp}(\omega) \subset U$ *is then*

$$\int_{\mathcal{M}} \omega = \int_V \psi^* \omega.$$

The neighborhood U is not required to be a sub-manifold of $\mathcal{M}$, although in applications $\varphi : V \subset \mathbb{R}^n \to U \subset \mathcal{M}$ will usually have full rank almost everywhere. Analogous to Def. 2.162, we can globalize Def. 2.164 using a partition of unity and multiple maps $\psi_i : V_i \subset \mathbb{R}^k \to U_i \subset \mathcal{M}$. The following two remarks exemplify Def. 2.164 and they recover the classical formulas for the integral of a vector field along a curve and over a surface.
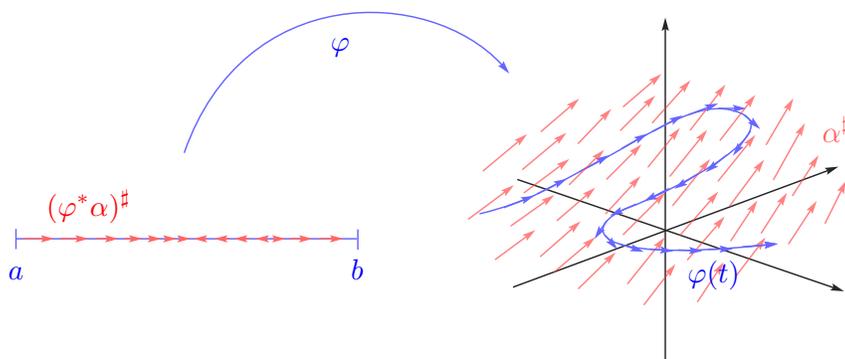
Figure 2.38: Integral of a 1-form $\alpha \in \Omega^1(\mathbb{R}^3)$ along a curve $\varphi : [a, b] \to \mathbb{R}^3$ in $\mathbb{R}^3$. What is integrated is the pullback $\varphi^*\alpha$ in the chart $[a, b]$.

Remark 2.113. Let $\gamma = (x^1(t), x^2(t), x^3(t)) : [a, b] \to \mathbb{R}^3$ be a curve in $\mathbb{R}^3$ and $\alpha \in \Omega^1(\mathbb{R}^3)$ be a 1-form field. To integrate α along γ we have to pullback the restriction $\bar{\alpha} = \alpha \mid_\gamma$ from $\mathbb{R}^3$ onto $[a, b] \subset \mathbb{R}$, cf. Fig. 2.38. Using the definition of the integral of a differential form in Def. 2.164 and the definition of the pullback for differential forms, Def. 2.139, it follows that

$$\int_\gamma \bar{\alpha} = \int_a^b (\gamma^*\bar{\alpha}) = \int_a^b \bar{\alpha}_{x(t)} \left(\left. \frac{\partial}{\partial x} \right|_{x(t)} \right) dt \tag{2.164}$$

where $\partial/\partial x \mid_{x(t)} = (T_t\gamma)e_1(t)$ is the image of the tangent vector $e_1(t)$ on the real line at $t \in [a, b]$ under the curve γ. In coordinates, for the standard basis in $\mathbb{R}^3$, we have $\bar{\alpha} = \alpha_1 e^1 + \alpha_2 e^2 + \alpha_3 e^3$ and $\partial/\partial x^i = X^1 e_1 + X^2 e_2 + X^3 e_3$ with $X^i = dx^i(t)/dt$, and the above pairing is hence

$$\bar{\alpha}_{x(t)} \left(\left. \frac{\partial}{\partial x} \right|_{x(t)} \right) = \alpha_1 X^1 e^1(e_1) + \alpha_2 X^2 e^2(e_2) + \alpha_3 X^3 e^3(e_3) = \alpha_i X^i \tag{2.165}$$

where the α_i and X^i are evaluated at $x(t)$. The integral is thus

$$\int_\gamma \bar{\alpha} = \int_a^b \alpha_i X^i dt \tag{2.166}$$

When one identifies the 1-form α with a vector using the musical isomorphism by $A = \alpha^\sharp$, which, as we remarked before, preserves in $\mathbb{R}^3$ with the standard basis the numerical value of the basis coefficients, then the above integral can be written as

$$\int_\gamma \bar{\alpha} = \int_a^b A \cdot X \, dt \tag{2.167}$$

which is the usual expression for the line integral of a vector field from vector calculus. It has to be emphasized that whenever line integrals of vector fields

are considered in applications, the vector field is given by $\alpha^\sharp$ and the quantity of interest is the 1-form field α.

Example 2.76. The change in kinetic energy for a particle travelling in a force field $\vec{F}$ along a space curve $\gamma = x(t) : [a, b] \to \mathbb{R}^3$ is classically given by

$$\Delta E = \int_\gamma \vec{F}(x(t)) \cdot X(x(t)) \, dt \tag{2.168}$$

where $X(x(t)) = dx(t)/dt$ is the velocity vector of the particle at $x(t)$. However, when one requires the above integral to be independent under a change of coordinates then a force vector field has to transform covariantly under a change of coordinates. By Remark 2.66, the contravariant vector X transforms as $X^i A_i^j$ under the linear isomorphism $A : \mathbb{R}^3 \to \mathbb{R}^3$ but for the value of the pairing $\vec{F} \cdot X$ to be preserved under the mapping $\vec{F}$ has to transform as $B_j^k F_k$, where B_j^k is the matrix inverse of A_i^j so that $A_i^j B_j^k = \delta_i^k$, since only then $X^i A_i^j B_j^k F_k = X^i \delta_i^k F_k = X^i F_i$. By the tensorality condition, an object that transforms covariantly is covariant and hence the force is a covector field or a differential 1-form field $f \in \Omega^1(\mathbb{R}^3) \cong T^*\mathcal{M}$. The change in the kinetic energy is then given by

$$\Delta E = \int_\gamma f = \int_a^b \gamma^* f \tag{2.169}$$

which agrees with the classical form by Remark 2.113.

Remark 2.114. Let $\mathcal{M} \subset \mathbb{R}^3$ be a surface in $\mathbb{R}^3$ and $\beta \in \Omega^2(\mathbb{R}^3)$ be a 2-form, and assume for simplicity that an atlas for $\mathcal{M}$ is given by a single chart (U, φ). To integrate β over $\mathcal{M}$ we have to pullback the restriction $\bar{\beta} = \alpha \mid_\mathcal{M}$ of β from $\mathbb{R}^3$ onto $\varphi(U)$. Using the definition of the integral of a differential form in Def. 2.164 and the definition of the pullback for differential forms, Def. 2.139, it follows that

$$\int_\mathcal{M} \beta = \int_{\varphi(U)} (\varphi^{-1})^* \beta = \int_{\varphi(U)} \beta \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) dx \, dy \tag{2.170}$$

where the $\partial/\partial x^i$ are the local basis for $T\mathcal{M}$. Using the representation of β and $\partial/\partial x, \partial/\partial y$ with respect to the canonical basis for $\mathbb{R}^3$,

$$\beta = \beta_1(e^2 \wedge e^3) + \beta_2(e^3 \wedge e^1) + \beta_3(e^1 \wedge e^2), \tag{2.171}$$

for the 2-form β, and

$$\frac{\partial}{\partial x} = X^1 e_1 + X^2 e_2 + X^3 e_3 \tag{2.172a}$$

$$\frac{\partial}{\partial y} = Y^1 e_1 + Y^2 e_2 + Y^3 e_3, \tag{2.172b}$$

for the local basis vectors, we obtain, using $(\alpha \wedge \beta)(v_1, v_2) = \alpha(v_1)\beta(v_2) - \alpha(v_2)\beta(v_1)$, for each of the basis function $e^i \wedge e^j$ that

$$
\left(e^i \wedge e^j\right)\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = e^i(X^1 e_1 + X^2 e_2 + X^3 e_3)\, e^j(Y^1 e_1 + Y^2 e_2 + Y^3 e_3)
$$
$$
- e^i(Y^1 e_1 + Y^2 e_2 + Y^3 e_3)\, e^j(X^1 e_1 + X^2 e_2 + X^3 e_3).
$$

With the biorthogonality $e^i(e_j) = \delta^i_j$ of the basis functions this is equivalent to

$$
\left(e^i \wedge e^j\right)\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = X^i e^i(e_i)\, Y^j e^j(e_j) - Y^i e^i(e_i)\, X^j e^j(e_j)
$$
$$
= X^i Y^j - Y^i X^j.
$$

The pullback therefore is

$$
\beta\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \begin{aligned} &\beta_1(X^2 Y^3 - Y^2 X^3) \\ &+ \beta_2(X^3 Y^1 - Y^1 X^3) \\ &+ \beta_3(X^1 Y^2 - Y^1 X^2). \end{aligned} \tag{2.173}
$$

The terms $X^i Y^j - Y^i X^j$ can be identified with the components of the cross product, and since these are obtained from the basis vectors $\partial/\partial x$ and $\partial/\partial y$ of the tangent space they can be identified with the normal vector $n = n^1 e_1 + n^2 e_2 + n^3 e_3$.[114] For the integral we hence obtain

$$
\int_{\mathcal{M}} \beta = \int_{\varphi(U)} \beta_i\, n^i\, dx\, dy \tag{2.174}
$$

Writing the normal as $n = \|n\|(\bar{n}^1 e_1 + \bar{n}^2 e_2 + \bar{n}^3 e_3)$, so that the components $\bar{n}^i$ form a normalized vector $\vec{n}$, and using the classical area differential dA that is defined as $dA = \|n\|\, dx\, dy$ we obtain for the integral

$$
\int_{\mathcal{M}} \beta = \int_{\varphi(U)} \beta_i\, \bar{n}^i\, dA. \tag{2.175}
$$

When the 2-form β is identified with a vector $B = (\star\beta)^\sharp$, and remembering that the Hodge dual and the musical isomorphism preserve the numerical value of the components in $\mathbb{R}^3$ with the standard basis, we have, with the dot product instead of the implicit sum,

$$
\int_{\mathcal{M}} \beta = \int_{\varphi(U)} B \cdot \vec{n}\, dA \tag{2.176}
$$

which is the usual expression for the surface or flux integral.

---

[114]It is important to note that the normal is *not* a vector but a covector that is only defined for $n-1$ sub-manifold of an n-dimensional volume manifolds (M, μ) by $\mu(\cdot, \partial/\partial u^1, \ldots \partial/\partial u^{n-1})$ where the $\partial/\partial u^i$ are the tangent vectors of the embedded manifold. The difference is for example important when a change of coordinates is considered.

Example 2.77. Gauss law for magnetism is given in classical notation by

$$\int_{\partial V} \vec{B} \cdot \vec{n} \, dA = 0$$

where $V \subset \mathbb{R}^3$ is a volume and $\vec{B}$ is the magnetic field. In the differential forms approach to electromagnetic theory[115] the magnetic field is a 2-form β and Gauss law takes the form $\int_{\partial V} \beta$ that is intrinsically invariant under coordinate transformations.

Remark 2.115. In applied fields one often employs Leibniz notation for the differential to express the infinitesimal change of a function $f(x)$ along x, for example

$$\frac{df(x)}{dx} = g(x). \tag{2.177}$$

A finite change is then obtained by "multiplying" by dx to obtain

$$df(x) = g(x) \, dx \tag{2.178}$$

and then integrating

$$\Delta f(x) = \int g(x) \, dx. \tag{2.179}$$

From our point of view one always has a differential form

$$\mathrm{d}f(x) = \frac{\partial f(x)}{\partial x} \, \mathrm{d}x = g(x) \, dx. \tag{2.180}$$

The "multiplication" in the classical approach using Leibniz notation is hence only needed to recover something that one forgot in the first place.

The following theorem is central to continuum theories and will be of considerable importance in the sequel.

Theorem 2.23 (Transport Theorem). *Let $(\mathcal{M}, \mu)$ be a volume manifold and $X \in \mathfrak{X}(\mathcal{M})$ a vector field on $\mathcal{M}$ with flow F_t. For $\omega_t = \Omega^k(\mathcal{M})$ a time-dependent k-form on $\mathcal{M}$ and $V \subset$ a k-dimensional, open sub-manifolds, it then holds that*

$$\frac{d}{dt} \int_{F_t(V)} \omega_t = \int_{F_t(V)} \left(\frac{\partial \omega_t}{\partial t} + \pounds_X \omega_t \right).$$

---

[115]Electromagnetic theory was one of the first areas where the importance of differential forms for computations was realized. Pioneering work on the subject was done by Deschamps ("Electromagnetics and Differential Forms") and a recent monograph on the subject is those by Bossavit (*Computational Electromagnetism: Variational Formulations, Complementarity, Edge Elements*). For the beautiful structure of Maxwell's equations in spacetime, and its relevance for computations, see the article by Stern et al. ("Geometric Computational Electrodynamics with Variational Integrators and Discrete Differential Forms").

When $\omega_t = f_t\,\bar{\mu}$, where $f_t \in \mathcal{F}(\mathcal{M})$ is a function and $\bar{\mu}$ a time invariant reference volume form, then

$$\frac{d}{dt}\int_{F_t(V)} f_t\,\mu = \int_{F_t(V)}\left(\frac{\partial f_t}{\partial t} + \operatorname{div}_\mu(f_t\,X)\right)\mu.$$

The equivalence of the two forms of the transport theorem follows from Proposition 2.76. The second form of the transport theorem is the one usually relevant in classical mechanics where the background space $\mathbb{R}^3$ is assumed to be absolute and hence an invariant volume form given by $dx^3 = dx \wedge dy \wedge dz$ exists.

Remark 2.116. The transport theorem often arises when $\omega_t \in \Omega^n(\mathcal{M})$ is a volume form and the quantity infinitesimally represented by ω_t is conserved. The left hand side of the transport equation then vanishes, expressing the invariance. For example, in ideal fluid dynamics one has conservation of mass so that for the fluid density $\rho(x,t) \in \Omega^3(B)$ one has

$$\frac{d}{dt}\int_{F_t(U)} \rho(x,t) = 0$$

where $U \subset B$ is an arbitrary subset of the fluid flow domain $B \subset \mathbb{R}^3$ and $F_t : \mathbb{R} \times B \to B$ is the flow generated by the fluid velocity field $u \in \mathfrak{X}(B)$. Since U is arbitrary and assuming B is smooth with smooth boundary ∂B we obtain from the right hand side of the transport equation in Theorem 2.23 that

$$\frac{\partial \rho}{\partial t} + \pounds_u \rho = 0$$

Using the second form of the transport theorem and omitting the invariant reference measure this is equivalent to

$$\frac{\partial \rho}{\partial t} + \operatorname{div}_\mu \bar{\rho} X = 0$$

which is the continuity equation of fluid dynamics.

Remark 2.117. The **material derivative**[116] can also be regarded as a special case of the transport theorem. Let $F_t : \mathbb{R} \times \mathcal{M} \to \mathcal{M}$ be a diffeomorphism of $\mathcal{M}$ that is generated by a vector field $X \in \mathfrak{X}(\mathcal{M})$, and let $\omega_t \in \Omega^k(\mathcal{M})$ be a time-dependent differential k-form on $\mathcal{M}$. Using the change of variable theorem,

$$\int_{F_t(W)} \omega_t = \int_W F_t^*\omega_t \tag{2.181a}$$

for a sufficiently well behaved subset $W \subset \mathcal{M}$, the transport theorem,

$$\int_{F_t(W)} f_t\mu = \int_{F_t(W)}\left(\frac{\partial f_t}{\partial t} + \operatorname{div}(f_tX)\right)\mu, \tag{2.181b}$$

---

[116]Chorin and Marsden, *A Mathematical Introduction to Fluid Mechanics*, p. 5.

becomes

$$F_t^* \left(\frac{d}{dt} F_t(x) \right) = F_t^* \left(\frac{\partial f_t}{\partial t} + \mathrm{div}(f_t X) \right). \tag{2.181c}$$

In $\mathbb{R}^3$, this can be written as

$$F_t^* \left(\frac{d}{dt} F_t(x) \right) = F_t^* \left(\frac{\partial f_t}{\partial t} + X \cdot \nabla f_t + f_t \, \mathrm{div}(X) \right) \tag{2.181d}$$

$$= F_t^* \left(\frac{\partial f_t}{\partial t} + X \cdot \nabla f_t \right) \tag{2.181e}$$

since

$$\varphi^*(f_t, \mathrm{div}(X)) = \varphi^*(f_t, \pounds_X \mu) = (\varphi^* f_t)(\varphi^* \pounds_X \mu) = 0 \tag{2.181f}$$

where $\varphi^* \pounds_X \mu$ vanishes since μ was assumed to be a volume form that is invariant under the flow. One therefore has

$$\frac{d}{dt} f_t = \frac{\partial f_t}{\partial t} + X \cdot \nabla f_t \tag{2.182}$$

which is the material derivative. The result holds for arbitrary domains, when ∇-notation is avoided, as long as an invariant volume form is available.

Using integration we can introduce an L_2-like inner product for differential forms on a manifold $\mathcal{M}$ using the metric pairing $\langle\!\langle \, , \rangle\!\rangle : \Omega^k(\mathcal{M}) \times \Omega^k(\mathcal{M}) \to \mathcal{F}(\mathcal{M})$ that was introduced in Def. 2.134 and generalized to manifolds in Def. 2.148.

Definition 2.165. *Let $(\mathcal{M}, g)$ be a oriented, Riemannian manifold with g-volume μ. The **inner product***

$$\langle\!\langle \, , \rangle\!\rangle : \Omega^k(\mathcal{M}) \times \Omega^k(\mathcal{M}) \to \mathbb{R}$$

of $\alpha, \beta \in \Omega^k(\mathcal{M})$ is

$$\langle\!\langle \alpha, \beta \rangle\!\rangle = \int_{\mathcal{M}} \alpha \wedge \star \beta.$$

Remark 2.118. Let $(\mathcal{M}, g)$ be a oriented, Riemannian manifold with g-volume μ, and $\alpha, \beta \in \Omega^k(\mathcal{M})$ have components α_i and β_i. Then

$$\langle\!\langle \alpha, \beta \rangle\!\rangle = \int_{\mathcal{M}} g^{ij} \alpha_i \, \beta_j = \int_{\mathcal{M}} \alpha_i \, \beta^i$$

Note that the components α_i and β_i are defined locally in a chart and the local expressions are "stitched together" using a partition of unity.

Example 2.78. Let $\varphi : \mathcal{M} \to \mathcal{M}$ be a diffeomorphism of a Riemannian manifold $(\mathcal{M}, g)$ and $\alpha, \beta \in \Omega^k(\mathcal{M})$ be k-forms on $\mathcal{M}$. Then

$$\langle\!\langle \alpha, \beta \rangle\!\rangle_g = \langle\!\langle \varphi^* \alpha, \varphi^* \beta \rangle\!\rangle_{\varphi^* g}$$

The pullback $\varphi^* g$ of the metric along the flow is known as Cauchy-Green stress tensor and it plays a central role in elasticity and fluid dynamics.

Stokes' Theorem Stokes' theorem will connect the integral of an n-form over an n-manifold $\mathcal{M}$ with boundary $\partial\mathcal{M}$ with the value of the integral of an associated $(n-1)$-form over $\partial\mathcal{M}$. Before we can state the theorem, however, we will have to introduce a suitable notion of orientation on the boundary $\partial\mathcal{M}$. When we consider the situation in the chart, then the functional $\lambda \in \mathbf{E}^*$ that defines the half-space over which the manifold with boundary is modelled by Def. 2.84 and Def. 2.85, induces two classes of vectors $v_{in}, v_{out} \in \mathbf{E}$ by $\lambda(v_{in}) > 0$ and $\lambda(v_{out}) < 0$ and these classes are independent of the chart.

Definition 2.166. *Let $\mathcal{M}$ be an oriented n-manifold with boundary $\partial\mathcal{M}$, and (U, φ) with $\varphi : U \to \mathbb{R}^n_\lambda$ be a positively oriented chart for $\mathcal{M}$. At $x \in \partial\mathcal{M}$ a basis $\{v_1, \ldots, v_{n-1}\}$ for $T_x(\partial\mathcal{M})$ is positively oriented if*

$$\{(T_x\varphi)^{-1}v_{out}, v_1, \ldots, v_{n-1}\}$$

is positively oriented in the orientation of $\mathcal{M}$, where $\lambda(v_{out}) < 0$ at $\varphi(x) \in \mathbb{R}^n_\lambda$.

Theorem 2.24 (Stokes Theorem). *Let $\mathcal{M}$ be an oriented, paracompact n-manifold with boundary $\partial\mathcal{M}$, and let $i : \partial\mathcal{M} \to \mathcal{M}$ be the inclusion map for the boundary. Then for an $(n-1)$-form $\alpha \in \Omega^{n-1}(\mathcal{M})$ with compact support*

$$\int_{\partial\mathcal{M}} i^*\alpha = \int_{\mathcal{M}} \mathrm{d}\alpha.$$

and $\int_{\mathcal{M}} \mathrm{d}\alpha = 0$ when $\partial\mathcal{M} = \emptyset$.

Stokes' theorem can be generalized to non-orientable manifolds using densities. The main difficulty thereby is to generalize the exterior derivative to this setting, and this leads to the concept of a twisted k-form. The theorem can also be extended to manifolds with piecewise smooth boundary.

Remark 2.119. Stokes' theorem assumes a paracompact manifold $\mathcal{M}$, that is $\mathcal{M}$ is required to be Hausdorff and every local covering of $\mathcal{M}$ has to have a locally finite refinement. A more practical criterion for a space to be paracompact is a result due to Stone[117] that shows that every metric space is paracompact.

For our applications we *do* need Stokes' theorem in full generality, for paracompact $\mathcal{M}$ and with $i^*\alpha$ on the left hand side; most treatments in the literature make stronger assumptions. The theorems of Gauss, Green, and Stokes in vector calculus are special cases of Stokes' theorem, as shown in the next example.

Example 2.79. Let $U \subset \mathbb{R}^3$ be a compact subset of $\mathbb{R}^3$ with boundary surface ∂U and let $\beta \in \Omega^2(U)$ be a 2-form on U. Stokes' theorem then reads

$$\int_U \mathrm{d}\beta = \int_{\partial U} \beta. \tag{2.183}$$

---

[117]Stone, "Paracompactness and product spaces".

From Example 2.71 we know that $\mathrm{d}\beta = \mathrm{div}(B)dx^3$ where $B = (\star\beta)^\sharp$ is the vector associated with the 2-form β. Moreover, in Remark 2.114 we showed that the integral of a 2-form over a surface in $\mathbb{R}^3$ can be written as the usual surface integral. Hence, Stokes' theorem takes the form

$$\int_U \mathrm{div}(B)\, dx^3 = \int_{\partial U} B \cdot \bar{n}\, dA \tag{2.184}$$

which is Gauss theorem.

Example 2.80. Let $\mathcal{M}$ be a $2n+1$ dimensional manifold and $\alpha, \beta \in \Omega^n\mathcal{M}$ and consider the integral

$$\int_{\mathcal{M}} \mathrm{d}\alpha \wedge \beta. \tag{2.185}$$

By the Leibniz rule, $\mathrm{d}(\alpha \wedge \beta) = \mathrm{d}\alpha \wedge \beta \pm \alpha \wedge \mathrm{d}\beta$, this equals

$$\int_{\mathcal{M}} \mathrm{d}\alpha \wedge \beta = \int_{\mathcal{M}} \mathrm{d}(\alpha \wedge \beta) \mp \int_{\mathcal{M}} \alpha \wedge \mathrm{d}\beta$$

and applying Stokes' theorem, Theorem 2.24, to the first term we obtain

$$\int_{\mathcal{M}} \mathrm{d}\alpha \wedge \beta = \int_{\partial\mathcal{M}} \alpha \wedge \beta \mp \int_{\mathcal{M}} \alpha \wedge \mathrm{d}\beta$$

which is an integration by parts formula for arbitrary differential forms and manifolds. If $\mathcal{M} = [a, b] \subset \mathbb{R}$ then the boundary is $\partial\mathcal{M} = \{a, b\}$ and the known formula for the real line is recovered.

Remark 2.120. Frequently, we will have to work with non-compact manifolds, for example the cotangent bundle T^*Q is non-compact even if $Q \subset \mathbb{R}^3$ is a compact subset of $\mathbb{R}^3$. In this case we will impose suitable decay conditions for the functions and densities of interest so that the integrals are well defined, and boundary terms vanish when integration by parts is used. The spaces of relevance are then $\mathcal{F}_c(Q)$ and $\mathrm{Den}_c(Q)$ where the 'c' standards for compact support.

A useful result that follows from Stokes' theorem is the following.

Corollary 2.13. *Let $\mathcal{M}$ be a compact manifold with boundary $\partial\mathcal{M}$, $X \in \mathfrak{X}(\mathcal{M})$ be a vector field on $\mathcal{M}$, and $\alpha \in \Omega^k(\mathcal{M})$ and $\beta \in \Omega^{n-k}(\mathcal{M})$. Then*

$$\int_{\mathcal{M}} \pounds_X\alpha \wedge \beta = -\int_{\mathcal{M}} \alpha \wedge \pounds_X\beta.$$

Remark 2.121. Many transport processes are described using the flux through a surface. With Stokes' theorem we can rewrite the transport theorem in

Theorem 2.23 as a flow through the surface of a volume. Let $\mathcal{M}$ be an n-manifold and $V \subset \mathcal{M}$ be a well behaved subset, and let ω_t be a time-dependent volume form. The transport theorem is

$$\frac{d}{dt} \int_{F_t(V)} \omega_t = \int_{F_t(V)} \left(\frac{\partial \omega_t}{\partial t} + \mathcal{L}_X \omega_t \right).$$

By Cartan's formula we have for the second term on the right hand side

$$\mathcal{L}_X \omega_t = \mathrm{d} \mathrm{i}_X \omega_t + \mathrm{i}_X \mathrm{d} \omega_t = \mathrm{d} \mathrm{i}_X \omega_t$$

since ω_t is a volume form and hence $\mathrm{d}\omega_t = 0$. The transport theorem can thus be written as

$$\frac{d}{dt} \int_{F_t(V)} \omega_t = \int_{F_t(V)} \frac{\partial \omega_t}{\partial t} + \int_{F_t(V)} \mathrm{d} \mathrm{i}_X \omega_t.$$

Using Stokes' theorem with the second term we obtain

$$\frac{d}{dt} \int_{F_t(V)} \omega_t = \int_{F_t(V)} \frac{\partial \omega_t}{\partial t} + \int_{\partial F_t(V)} \mathrm{i}_X \omega_t.$$

The integral over the boundary is evaluated using the tangent vectors $\partial/\partial x^i$ of the boundary $\partial F_t(V)$ as arguments to $\mathrm{i}_X \omega_t$, cf. Remark 2.114. But since for any differential form α one has $\alpha(\dots, X, \dots, X, \dots) = 0$, the components of X that lie in the tangent space of $\partial F_t(V)$ are in the kernel of the 1-form $\omega_t(\cdot, \partial/\partial x^i, \dots, \partial/\partial x^n)$. Hence, $\mathrm{i}_X \omega_t$ corresponds to the usual notion of the normal. For example, as discussed in Remark 2.107, in $\mathbb{R}^3$ with the standard volume form dx^3 a coordinate calculation shows that

$$\mathrm{i}_X dx^3 = X_1 (dx^2 \wedge dx^3) - X_2 (dx^1 \wedge dx^3) + X_3 (dx^1 \wedge dx^2)$$

where the vector field $X \in \mathfrak{X}(\mathbb{R}^3)$ is $X = X^1 e_1 + X^2 e_2 + X^3 e_3$ and numerically $X_1 = X^1$. With Remark 2.114, the integral over the boundary hence indeed becomes the usual flux integral where the normal component of X determines the transport out of a test volume. In the literature, heuristic derivations often start by postulating a flux through the surface and then using Gauss' theorem to obtain a volume integral that leads to a form of the transport equation.

Riemannian Metrics Revisited With integration on manifolds now being well defined, we can complete our discussion of Riemannian metrics and the structure they induce.

Definition 2.167. *Let $(\mathcal{M}, g)$ be a Riemannian manifold, and let $\gamma : [a, b] \to \mathcal{M}$ be a curve on $\mathcal{M}$. The **length** $\ell(\gamma)$ **of** γ is*

$$\ell(\gamma) = \int_a^b \|\gamma'(t)\| \, dt$$

and the **energy of** γ is

$$E(\gamma) = \int_a^b \|\gamma'(t)\|^2 \, dt$$

with $\gamma'(t) = d\gamma(t)/dt$.

In the physics literature in Lagrangian mechanics the energy is also known as the action along the path, cf. Example 2.32. It is also easy to see that the length of a curve, but *not* its energy, is invariant under a reparametrization of t.

Proposition 2.77. *Let $(\mathcal{M}, g)$ be a Riemannian manifold, and let $\gamma : [a,b] \to \mathcal{M}$ be a curve on $\mathcal{M}$. Then*

$$\ell(\gamma)^2 \leq 2 \, (b-a) \, E(\gamma)$$

and equality hold if and only if $\gamma'(t)$ is constant.

The above proposition is a consequence of the Cauchy-Schwarz inequality.

Definition 2.168. *Let $(\mathcal{M}, g)$ be a Riemannian manifold. The shortest path between two points on $\mathcal{M}$ is the **geodesic** between the points.*

It follows that a geodesic is a stationary point of the length functional $\ell(\gamma)$ and it can hence locally be computed using an Euler-Lagrange-type equation known as the geodesic equation. In finite dimensions this leads to Christoffel symbols, while in infinite dimensions a geodesic strays have to be employed. By Proposition 2.77 it is possible to find geodesics using the energy of a curve and this is often the ansatz used in practice. Using geodesics one can define the exponential map $\exp_m(\mathcal{M})$ of $\mathcal{M}$ at a point $m \in \mathcal{M}$ which provides a diffeomorphism between a neighborhood of $0 \in T_m\mathcal{M}$ and $m \in \mathcal{M}$.

Remark 2.122. The length and energy of a curve introduced in Def. 2.167, and hence also geodesics, only depend on the norm induced by the metric. Hence, what is needed in each tangent space is not necessarily an inner product. This insight leads to the notion of a Finsler metric and a manifold with such a metric is known as a Finsler manifold.

Hodge Decomposition   An application that combines a variety of the concepts introduced previously is the Hodge decomposition. Recall that a differential k-form α is closed if $d\alpha = 0$ and exact if $d\beta = \alpha$ for some $(k-1)$-form β. The codifferential $\delta : \Omega^{k+1}(\mathcal{M}) \to \Omega^k(\mathcal{M})$ on a n-dimensional Riemannian manifold $(\mathcal{M}, g)$ was $\delta\beta = \pm \star d \star \beta$ where $\beta \in \Omega^k(\mathcal{M})$ and the sign depends on n, k, and the index of the metric. A differential k-form β is co-closed if $\delta\beta = 0$. Finally, the Laplace-de Rahm operator $\Delta : \Omega^k(\mathcal{M}) \to \Omega^k(\mathcal{M})$ on a Riemannian manifold is $\Delta = d\delta + \delta d$, and the elements of the kernel $\mathcal{H}^k$ of the Laplace-de Rahm operator Δ are known as harmonic forms.

Remark 2.123. Using Stokes' theorem one can show that on a Riemannian manifold $(\mathcal{M}, g)$ *without* boundary for $\alpha \in \Omega^{k-1}(\mathcal{M})$ and $\beta \in \Omega^k(\mathcal{M})$ one has

$$\langle\!\langle \mathrm{d}\alpha, \beta \rangle\!\rangle = \langle\!\langle \alpha, \delta\beta \rangle\!\rangle$$

and the exterior derivative d and the codifferential δ are adjoint operators under the inner product for k-forms introduced in Def. 2.165. The result implies for $\alpha, \beta \in \Omega^k(\mathcal{M})$ that

$$\langle\!\langle \Delta\alpha, \beta \rangle\!\rangle = \langle\!\langle \mathrm{d}\delta + \delta\mathrm{d}\alpha, \beta \rangle\!\rangle = \langle\!\langle \mathrm{d}\delta\alpha, \beta \rangle\!\rangle + \langle\!\langle \delta\mathrm{d}\alpha, \beta \rangle\!\rangle$$

$$= \langle\!\langle \alpha, \mathrm{d}\delta\beta \rangle\!\rangle + \langle\!\langle \alpha, \delta\mathrm{d}\beta \rangle\!\rangle = \langle\!\langle \alpha, \mathrm{d}\delta + \delta\mathrm{d}\beta \rangle\!\rangle = \langle\!\langle \alpha, \Delta\beta \rangle\!\rangle$$

and the Laplace-de Rahm operator is self-adjoint and positive definite.

The Hodge decomposition theorem can be seen as a generalization of the Helmholtz decomposition of vector fields that plays an important role in fluid mechanics and electromagnetism—as usual, the vector fields in the Helmholtz decomposition are in fact differential forms that are identified with vectors using the standard metric on $\mathbb{R}^3$.

Theorem 2.25 (Hodge decomposition). *Let $\mathcal{M}$ be a compact, oriented Riemannian manifold $(\mathcal{M}, g)$ without boundary. Any $\omega \in \Omega^k(\mathcal{M})$ admits the $\langle\!\langle \, , \, \rangle\!\rangle$-orthogonal decomposition*

$$\omega = \mathrm{d}\alpha + \delta\beta + \gamma$$

where $\gamma \in \mathcal{H}^k(\mathcal{M})$ is harmonic, and $\alpha \in \Omega^{k-1}(\mathcal{M})$, $\beta \in \Omega^{k+1}(\mathcal{M})$, and $\gamma \in \Omega^k(\mathcal{M})$.

The theorem hence asserts that on a Riemannian manifold every differential k-form can be decomposed into an exact form $\mathrm{d}\alpha$, a co-exact form $\delta\beta$, and a harmonic form γ.

Corollary 2.14. *Let $\mathcal{M}$ be a Riemannian manifold as in the foregoing theorem. Then the space of differential k-forms $\Omega^k(\mathcal{M})$ has an orthogonal decomposition*

$$\Omega^k(\mathcal{M}) = \mathrm{d}\Omega^k(\mathcal{M}) \otimes \delta\Omega^k(\mathcal{M}) \otimes \mathcal{H}^k(\mathcal{M}).$$

where $\mathrm{d}\Omega^k(\mathcal{M})$ is the space of exact k-forms on $\mathcal{M}$, $\delta\Omega^k(\mathcal{M})$ the space of co-exact k-forms, and $\mathcal{H}^k(\mathcal{M})$ is the space of harmonic forms.

Remark 2.124. The orthogonality of the components in the Hodge decomposition is an immediate consequence of $\mathrm{dd} = 0$ and $\delta\delta = 0$ and the adjointness of d and δ.

Definition 2.169. *Let $\mathcal{M}$ be a compact, oriented Riemannian manifold $(\mathcal{M}, g)$. The kth **de Rahm cohomology group** $\mathrm{H}^k(\mathcal{M})$ **of** $\mathcal{M}$ is the equivalence class*

$$\mathrm{H}^k(\mathcal{M}) = \frac{\ker\left(\mathrm{d}^k\right)}{\mathrm{ran}\left(\mathrm{d}^{k-1}\right)}$$

of closed differential k-forms in $\ker\left(\mathrm{d}^k\right)$ modulo exact differential k-forms in $\mathrm{ran}\left(\mathrm{d}^{k-1}\right)$, where d^k is the exterior derivative acting on the space of k-forms.

Two closed differential k-forms in $\ker\left(\mathrm{d}^k\right)$ are hence identified if they differ by an exact $(k-1)$-form in $\mathrm{ran}\left(\mathrm{d}^{k-1}\right)$. The following is a consequence of the Hodge decomposition.

Corollary 2.15. *Let $\mathcal{M}$ be a compact, oriented Riemannian manifold $(\mathcal{M}, g)$ without boundary. The space of harmonic k-forms $\mathcal{H}^k(\mathcal{M})$ on $\mathcal{M}$ and the kth de Rham cohomology group of $\mathcal{M}$ are isomorphic as vector spaces.*

The corollary for example implies that H^k is finite dimensional, since the kernel of any elliptic operator such as the Laplace-de Rham operator Δ is finite dimensional on a compact manifold. Care is required when the Hodge decomposition is extended to manifolds with boundary since the exterior derivative and the codifferential are in this case no longer adjoint.

Definition 2.170. *Let $(\mathcal{M}, g)$ be a compact, oriented Riemannian manifold with boundary $\partial\mathcal{M}$, and let $i : \partial\mathcal{M} \to \mathcal{M}$ be the inclusion map. A differential form $\alpha \in \Omega^k$ is **normal to** $\partial\mathcal{M}$ if $\partial\mathcal{M}$ if $i^*(\star\alpha) = 0$ and it is **tangential to** $\partial\mathcal{M}$ if $i^*\alpha = 0$.*

It is easy to check that these notions do agree with the usual ones for vectors using the Riemannian structure available on $\mathcal{M}$; that is $\alpha \in \Omega^1\mathcal{M}$ is tangent to $\partial\mathcal{M}$ if and only if $\alpha^\sharp$ is tangent to $\partial\mathcal{M}$, and $\alpha \in \Omega^{n-1}\mathcal{M}$ is normal to $\partial\mathcal{M}$ if and only if $(\star\alpha)^\sharp$ is normal to $\partial\mathcal{M}$.

Theorem 2.26 (Hodge decomposition for manifolds with boundary)**.** *Let $(\mathcal{M}, g)$ be a compact, oriented Riemannian manifold with boundary $\partial\mathcal{M}$, and let*

$$\Omega_t^k(\mathcal{M}) = \left\{\alpha \in \Omega^k(\mathcal{M}) \mid \alpha \text{ is tangent to } \partial\mathcal{M}\right\}$$

$$\Omega_n^k(\mathcal{M}) = \left\{\alpha \in \Omega^k(\mathcal{M}) \mid \alpha \text{ is normal to } \partial\mathcal{M}\right\}$$

$$\mathcal{H}_n^k(\mathcal{M}) = \left\{\alpha \in \Omega^k(\mathcal{M}) \mid \mathrm{d}\alpha = 0, \delta\alpha = 0\right\}.$$

Then any differential k-form $\omega \in \Omega^k(\mathcal{M})$ admits the orthogonal decomposition

$$\omega = \mathrm{d}\alpha_t + \delta\beta_n + \gamma$$

for $\gamma \in \mathcal{H}^k(\mathcal{M})$ and $\alpha \in \Omega_t^{k-1}(\mathcal{M})$ and $\beta \in \Omega_n^{k+1}(\mathcal{M})$.

Corollary 2.16. *Let $\mathcal{M}$ be a Riemannian manifold as in the foregoing theorem. Then the space of differential k-forms $\Omega^k(\mathcal{M})$ has an orthogonal decomposition*

$$\Omega^k(\mathcal{M}) = \mathrm{d}\Omega_t^k(\mathcal{M}) \otimes \delta\Omega_n^k(\mathcal{M}) \otimes \mathcal{H}^k(\mathcal{M}).$$

where $\mathrm{d}\Omega_t^k(\mathcal{M})$ is the space of exact k-forms on $\mathcal{M}$ and tangent to $\partial\mathcal{M}$ and $\delta\Omega_n^k(\mathcal{M})$ the space of co-exact k-forms normal to $\partial\mathcal{M}$, and $\mathcal{H}^k(\mathcal{M})$ is the space of harmonic forms.

Remark 2.125. Despite their somewhat abstract appearance, the Hodge decomposition and the above corollaries have immediate physical applications, for example in fluid dynamics and in electromagnetism.

Example 2.81. Let $U \subset \mathbb{R}^3$ and assume the differential 1-form $\omega \in \Omega^1(U)$ has compact support on U. Then by the Hodge decomposition

$$\omega = \mathrm{d}f + \delta\beta + \gamma$$

for some function $f \in \mathcal{F}(U)$, some differential 2-form $\beta \in \Omega^2(U)$, and a harmonic form $\gamma \in \mathcal{H}^1(U)$. It follows that $\mathrm{dd}f = 0$ and $\delta\delta\beta = 0$, cf. Remark 2.102, and by identifying the differential forms with vectors this is equivalent to $\mathrm{curl}\,((\mathrm{d}f)^\sharp) = 0$ and $\mathrm{div}(\beta^\sharp) = 0$. Recognizing that the metric on $\mathbb{R}^3$ provides an isomorphism between $T\mathbb{R}^3$ and $T^*\mathbb{R}^3$ we have that any vector field $C = \omega^\sharp$ can be written as

$$C = F_l + F_t = \mathrm{grad}\,(f) + \mathrm{curl}\,(B)$$

where $B = (\delta\beta)^\sharp$ and we disregarded the harmonic term. The above formula is the Helmholtz decomposition of a vector field into a longitudinal part $F_l = \mathrm{grad}\,(f)$ that is irrotational and a transverse part $F_t = \mathrm{curl}\,(B)$ that is divergence free.

2.3.2.7 Flows on Function Spaces[118]

The action of a vector field $X \in \mathfrak{X}(\mathcal{M})$ with flow $F_t : \mathbb{R} \times \mathcal{M} \to \mathcal{M}$ on functions $f \in \mathcal{F}(\mathcal{M})$ is infinitesimally described by the Lie derivative $\mathcal{L}_X f$ and for finite times by the pullback $f_t = F_t^* f = f \circ F_t$. The description of such flows $t \to f_t \in \mathcal{F}(\mathcal{M})$ using linear operators will be discussed in this section: **We will replace a finite dimensional but nonlinear problem with an infinite dimensional but linear one**. This change in perspective will open up the rich toolbox of operator theory, and it is used for example when the long time behaviour of dynamical systems is of interest, such as in ergodic theory and stability theory.

Semi-groups Semi-groups of operators provide a finite time description of initial value problems, and they hence play an important role in the modern theory of dynamical systems.[119]

Definition 2.171. *Let $\mathcal{H}$ be a Hilbert space. A **one-parameter semi-group** $\{U(t) \mid t \in T^+\}$ **on $\mathcal{H}$ is the family of bounded operators $U(t) : \mathcal{H} \to \mathcal{H}$ with $T^+ \subseteq [0, \infty)$ such that*

i) $U(0) = \mathrm{id}$;

ii) $U(t + s) = U(t) \circ U(s), \ t, s \in T^+$ (semi-group property).

---

[118]The reader should refer to Chapter 2.2.2 for the notions of Hilbert space theory needed in this section.

[119]For an introduction to semi-group theory see for example the book by Engel and Nagel (*One-Parameter Semigroups for Linear Evolution Equations*).

*The family of operators $U(t)$ is a **discrete-time semi-group** if T^+ is discrete, and it is **strongly continuous** if for every $f \in \mathcal{H}$ the limit*

$$\lim_{t \to 0} U(t)f = f$$

*converges in the strong operator topology. The **infinitesimal generator** $A \in L(\mathcal{H}, \mathcal{H})$ of a one-parameter semi-group is*

$$Af = \frac{d}{dt}\left(U(t)f\right) = \lim_{t \to 0} \frac{U(t)f - f}{t}.$$

Remark 2.126. Arguably the simplest differential equation is

$$\frac{d}{dt}F(t) = aF(t) \tag{2.186}$$

where $a \in \mathbb{R}$ is a real-valued constant and $F : \mathbb{R} \to \mathbb{R}$. As is well known, the solution to the equation for all $t > 0$ is given by

$$F(t) = e^{ta} \tag{2.187}$$

and it is elementary that

$$F(0) = e^0 = 1 \tag{2.188a}$$

$$F(t + s) = e^{a(t+s)} = e^{at}\, e^{as}. \tag{2.188b}$$

The initial value problem in Eq. 2.186, whose infinitesimal rate of change is described by a, hence naturally leads to the semi-group $F(t) = e^{at}$. The connection between the infinitesimal description of initial value problems using differential equations and the finite time description using semi-groups can be traced back at least to Cauchy in the early 19th century.[120] The importance of both descriptions in the context of physics arises from determining the future state of a system at time $t > 0$ given the state at $t = 0$.

A one-parameter semi-group of operators where the semi-group property extends to all t forms a proper group.

Definition 2.172. *Let $\mathcal{H}$ be a Hilbert space. A **one-parameter group** $\{U(t) \mid t \in T\}$ **on** $\mathcal{H}$ is a one-parameter semi-group where the group property holds for all $t \in T = (-\infty, \infty)$.*

The group property will be discussed in more detail in Sec. 2.3.3 when we consider Lie groups in more detail.

---

[120]See Chapter I and Chapter VII in the book by Engel and Nagel (*One-Parameter Semigroups for Linear Evolution Equations*) for a more detailed discussion.

Stone's theorem   The directional derivative defined by a vector field $X \in \mathfrak{X}(\mathcal{M})$ provides a linear operator $A : \mathcal{F}(\mathcal{M}) \to \mathcal{F}(\mathcal{M})$ on the space of smooth functions $\mathcal{F}(\mathcal{M})$ on a manifold $\mathcal{M}$. When the operator A is anti-self-adjoint, satisfying $\langle Af, g \rangle = \langle f, -Ag \rangle$, then Stone's theorem characterizes the induced flow on $\mathcal{F}(\mathcal{M})$ as a subspace of a suitable Hilbert space such as $L_2(\mathcal{M})$.

Theorem 2.27 (Stone's theorem, real case).[121]   *Let $\mathcal{H}$ be a real Hilbert space and $A : \mathcal{H} \to \mathcal{H}$ be a anti-self-adjoint operator. Then A generates a one-parameter group $\{U(t) \mid t \in T\}$ of real unitary operators on $\mathcal{H}$ and if A is bounded then*

$$U(t) = e^{tA} = \mathrm{id} + tA + \frac{1}{2!}(tA)^2 + \dots$$

Conversely, every one parameter group of isometries has a anti-self-adjoint operator as infinitesimal generator.

In the literature, $U(t) = e^{tA}$ is often written for a one-parameter group even when A is not bounded, see also Remark 2.126.

Remark 2.127. For a complex Hilbert space, Stone's theorem holds when anti-self-adjointness of the infinitesimal generator A is replaced with self-adjointness. The importance of the theorem for dynamical systems follows from a result showing that a smooth, divergence-free vector field X on a volume manifold $(\mathcal{M}, \mu)$ with a complete flow defines an (essentially) self-adjoint operator $A = iX$ on the space C_c^∞ of functions with complex support in $L_2(\mathcal{M}, \mu)$. Conversely, when a vector field X is smooth and divergence-free and (essentially) self-adjoint as an operator iX then the flow associated with X is complete.

Remark 2.128. Stone's theorem is a special case of the Hille-Yosida theorem that provides a complete, albeit rather technical, characterization of the generators of strongly continuous semi-groups.

Remark 2.129. A result closely related to Stone's theorem is a lemma by Koopman[122], obtained at about the same time, which states, in modern terminology, that any volume preserving diffeomorphism g has an associated unitary operator U_g in L_2. Since volume preserving diffeomorphisms are the configuration space for ideal fluid dynamics, see the discussion in Sec. 2.3.3.3 and Example 2.147, this shows that **even nonlinear fluid dynamics has an equivalent description using linear operators** on an infinite dimensional function space.[123]

---

[121]Marsden and Hughes, *Mathematical Foundations of Elasticity*, Theorem 6.2.18.3.

[122] (Koopman, "Hamiltonian Systems and Transformations in Hilbert Space"), see also the often overlooked paper by von Neumann ("Zur Operatorenmethode In der Klassischen Mechanik") and (van Hove, "Sur le probléme des relations entre les transformations unitaires de la Mécanique quantique et les transformations canoniques de la Mécanique classique.", Sec. 15).

[123]Given the foregoing discussion it probably comes at no surprise that Koopman's lemma provides the only known ansatz to obtain a structure preserving discretization of ideal fluid dynamics (Mullen et al., "Energy-Preserving Integrators for Fluid Animation"; Pavlov et al., "Structure-preserving discretization of incompressible fluids").

Remark 2.130. Two areas where flows induced on function spaces play a central role are ergodic theory and quantum mechanics, and the above results were first obtained in this context. For the real case, usually minor modifications of the original statements in the complex setting are necessary, see for example Remark 2.127.

Definition 2.173. *Let $(\mathcal{M}, \mu)$ be a volume manifold. The space of half-densities on $\mathcal{M}$ carries a natural inner product and its completion is the **intrinsic Hilbert space $\mathfrak{H}(\mathcal{M})$ on $\mathcal{M}$**.*

It is not hard to see, keeping the non-degeneracy of the volume form in mind, that the intrinsic Hilbert space $\mathfrak{H}(\mathcal{M})$ is isomorphic to $L_2(\mathcal{M}, \mu)$ and that the isomorphism is given by $f \leftrightarrow f\sqrt{\mu}$.

2.3.3   Lie Groups

Groups provide the modern language to characterize symmetries: the invariance of a system under reversible changes. For example, permutations preserve the number and labels of elements while they allow to describe all different orderings and how to transition between them. Lie groups, which have the additional structure of a smooth manifold, enable to describe continuous symmetries, where the group elements are indexed using a continuous variable. Such symmetries are central to modern physics, and by Noether's theorem every continuous symmetry has an associated conserved quantity. For example, energy conservation is associated with invariance under the translation group on the time axis—for most physical systems absolute time is irrelevant—and momentum conservation is associated with the action of the three dimensional translation group in Euclidean space—usually, the absolute position of a physical system in space does not affect its time evolution.[124] The above symmetries can be found in most systems in classical mechanics, and arguably they do not require the complexities of group theory for their description. The simplest systems where a modern perspective is essential is the Euler top, the rigid body whose center of mass is fixed in space.[125] The system is invariant under the action of the Lie group SO(3), with angular momentum being the associated Noetherian quantity. However, the group SO(3) also provides the configuration space of the system, since every state of the body can be described by a rotation, and only when this is considered is an understanding of the equations of motion of the body possible. Analogously, the ideal Euler fluid has the Lie group $\mathrm{Diff}_\mu(\mathcal{M})$ of volume preserving diffeomorphisms as configurations space, and exploiting the Lie group structure is essential to gain insight into the system, for example into why vorticity is conserved along the fluid flow.

---

[124]In space-time, however, where there is only a single momentum 4-vector, there is only a single symmetry associated with space and time translation. As usual, in the regime of classical mechanics the splitting of space and time is a very accurate approximation.

[125]See (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*) for a modern discussion of the surprisingly rich theory of the rigid body.
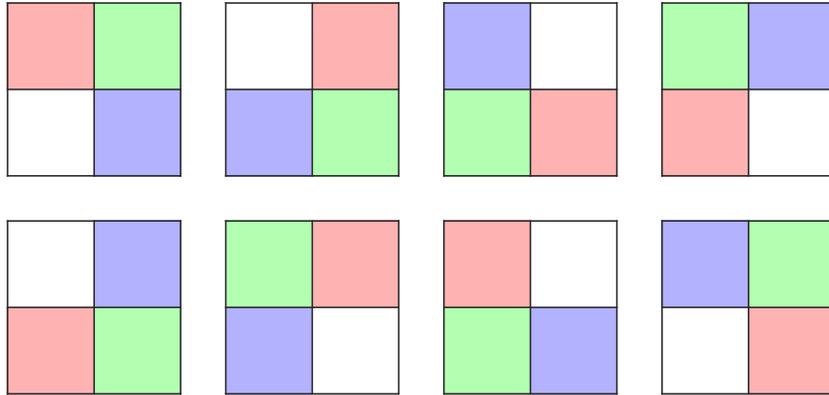
Figure 2.39: A group can be considered as the set of possible transformations of a set. A Lie group has the additional structure of a smooth manifold, that is the set of transformations is smooth and infinitely differentiable, so that for example the velocity of a smoothly varying transformation is defined.

In this section we will present an elementary introduction to Lie groups and their action on manifolds, which will be central for the description of physical systems with symmetry. In the first part of the section we will not be careful to distinguish a Lie group and its representation, but the difference will become clear in Sec. 2.3.3.2.

2.3.3.1 Definition and Properties[126]

Before introducing Lie groups, let us briefly recall the definition of a group from Chapter 2.2.

Definition 2.174. *A **group** G is a set with a binary **group multiplication***

$$g \circ h = gh : G \times G \to G, \quad g, h \in G$$

*that is closed in the set and associative so that $f \circ (g \circ h) = (f \circ g) \circ h$. The **identity element** e of a group is the unique element such that*

$$e \circ g = g \circ e = g$$

*for all $g \in G$. For every group element $g \in G$ there exists a unique **inverse element** g^{-1} such that*

$$g \circ g^{-1} = g^{-1} \circ g = e.$$

*A group is **Abelian** if group multiplication commutes and $g \circ h = h \circ g$ for all $g, h \in G$.*

---

[126]An elementary and largely self-contained introduction to Lie groups can be found in (Stillwell, *Naive Lie Theory*).

Remark 2.131. More intuitively, a group can be defined as a collection of transformation of a set such that every transformation has an inverse belonging to the collection and the composition of two transformation is another transformation in the collection, cf. Fig. 2.39.

Example 2.82. Permutations, as discussed in Remark 2.62, form a group.

A Lie group is a group with a smooth structure provided by an underlying manifold.

Definition 2.175. *A **Lie group** G is a manifold that has a group structure so that group multiplication*

$$\mu : G \times G \to G := g \circ h = gh$$

*is smooth as an operation on the manifold. The **inversion map** $\mathrm{Inv} : G \to G :$ $\mathrm{Inv}(g) \to g^{-1}$ of a Lie group is the smooth map that assigns to each element $g \in G$ its inverse $g^{-1} \in G$.*

Example 2.83. Consider the circle S^1. For two points $\theta_1, \theta_2 \in \mathbb{S}^2$ on the manifold group multiplication can be defined as

$$\theta_1 \circ \theta_2 = \theta_1 + \theta_2,$$

with addition modulo 2π, which is closed and smooth, with smoothness being apparent for example when written as rotation in the plane. The inversion map is then

$$\mathrm{Inv}(\theta_1) = -\theta_1,$$

which is again smooth, and the identity is

$$e = 0.$$

Hence S^1 with the above operations is a Lie group, known as the circle group.

Example 2.84. Every vector space V is a Lie group whose manifold structure is provided by the vector space structure of V. For all $a, b \in V$, group multiplication is vector addition,

$$a \circ b = a + b,$$

the inversion map is

$$\mathrm{Inv}(a) = -a,$$

and the identity is the zero vector,

$$e = 0.$$

When a vector space is regarded as a Lie group it is denoted as a **vector group**.

Example 2.85.[127] The open set of invertible isomorphisms of $\mathbb{R}^n$ in $L(\mathbb{R}^n, \mathbb{R}^n)$ forms the **general linear group** $\mathrm{GL}(n, \mathbb{R})$. That $\mathrm{GL}(n, \mathbb{R})$ has a smooth manifold structure can be seen by considering the determinant map det : $\mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}$ which is a smooth map. But then also the pre-image in $L(\mathbb{R}^n, \mathbb{R}^n)$ is a smooth subset of $L(\mathbb{R}^n, \mathbb{R}^n)$. Choosing a basis for $\mathbb{R}^n$, each element in $\mathrm{GL}(n, \mathbb{R})$ can be represented by an invertible matrix. For $A, B \in \mathrm{GL}(n, \mathbb{R})$, group multiplication is then matrix multiplication

$$A \circ B = A + B,$$

the inversion map is matrix inversion

$$\mathrm{Inv}(A) = A^{-1},$$

and the group identity is the identity matrix,

$$e = I.$$

The smoothness of the group operations follows immediately from the smoothness of the matrix operations in the entries.

Remark 2.132. Our definition of a Lie group does not make any assumptions on the underlying manifold and it can hence be infinite dimensional, for example a Banach manifold. However, for important examples of infinite dimensional Lie groups, such as diffeomorphism groups, much care is required and the above definition does not strictly apply, although it is customary in the literature to formally treat even these groups according to Def. 2.175 and assume everything is well behaved. We will consider these questions in more detail in Sec. 2.3.3.3.

Definition 2.176. *The **left and right translation maps** of a Lie group G are the maps*

$$L_g : G \to G : gh \qquad R_g : G \to G : hg.$$

Some useful properties of the translation maps are summarized in the following proposition.

Proposition 2.78. *Let G be a Lie group. The left and right translation maps L_g and R_g then satisfy*

i) *Composition: $L_g \circ L_h = L_{gh}$ and $R_g \circ R_h = R_{hg}$;*

ii) *Inversion: $(L_g)^{-1} = L_{g^{-1}}$ and $(R_g)^{-1} = R_{g^{-1}}$;*

iii) *Commutativity: $L_g \circ R_h = R_h \circ L_g$;*

iv) *Tangent map: $T_h L_g : T_h G \to T_{gh} G$ and $T_h R_g : T_h G \to T_{hg} G$ are isomorphisms of the tangent spaces of G.*

---

[127]The general linear group can be defined for arbitrary finite fields but we will restrict us, as usual, to the real case, that is $\mathrm{GL}((, \mathbb{R})n, \mathbb{R})$.

Remark 2.133. On a Lie group G, an atlas for the manifold structure of G can be obtained from a single admissible chart by group translation. For example, let $\{U, \varphi\}$ be a chart around the identity with $\varphi : U \to \mathbf{E}$. Then $\{U_g, \varphi_g\}$ defined by

$$U_g = L_g(U) = \{L_g h \mid h \in U\} \qquad \varphi_g = \varphi \circ L_{g^{-1}} : U_g \to \mathbf{E}$$

is a chart around $g \in G$. That the resulting transition maps are well behaved follows from the smoothness of group multiplication and inversion.

Invariant vector fields, which are introduced next, play a central role in the definition of the Lie algebra and for the description of dynamics on Lie groups. They are also a natural or distinguished class of vector fields since they commute with the translation action on the group.

Definition 2.177. *Let G be a Lie group. A vector field $X \in \mathfrak{X}(G)$ on G is* **left invariant** *when*

$$L_g^* X(h) = (T_h L_g)(X(h)) = X(L_g h) = X(g \circ h)$$

for all $g \in G$, and it is **right invariant** *when*

$$R_g^* X(h) = (T_h R_g)(X(h)) = X(R_g h) = X(h \circ g).$$

The spaces of left and right invariant vector fields are denoted by $\mathfrak{X}_L(G)$ and $\mathfrak{X}_R(G)$ respectively.

For invariant vector fields hence the following diagram holds,

$$
\begin{array}{ccc}
G & \xrightarrow{\ L_g/R_g\ } & G \\
X \downarrow & & \downarrow X \\
TG & \xrightarrow[\ TL_g/TR_g\ ]{} & TG
\end{array}
$$

Example 2.86. For a vector group the invariant vector fields are the constant ones.

Proposition 2.79. *Let $\xi \in T_e G$. Then*

$$X_\xi = (T_e L_g)\xi \qquad X_\xi = (T_e R_g)\xi$$

for all $g \in G$ define a left and right invariant vector field on G, respectively. Moreover, there is a one-to-one correspondence between elements in $T_e G$ and left and right invariant vector fields, and as vector spaces

$$T_e G \cong \mathfrak{X}_L(G) \qquad T_e G \cong \mathfrak{X}_R(G).$$

Intuitively, the above proposition states that for the study of left or right invariant vector fields on a Lie group G, it suffices to consider the tangent space T_eG at the identity. This motivates the definition of the Lie algebra of a Lie group which is T_eG together with a bilinear bracket. However, before we will consider Lie algebras associated with Lie groups we will introduce them as structures in their own right, and for example in the infinite dimensional case there do exist Lie algebras which do not arise from a Lie group.

Definition 2.178. *A **Lie algebra** $\mathfrak{g}$ is a vector space together with the **Lie bracket** $[\,,] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ satisfying*

i) bilinearity: $[a\,\xi + b\,\eta, \gamma] = a\,[\xi, \gamma] + b\,[\eta, \gamma]$ *and* $[\xi\,, c\,\eta + d\,\gamma] = c\,[\xi, \eta] + d\,[\xi, \gamma]$

ii) anti-symmetry: $[\xi, \eta] = -[\eta, \xi]$ *(or equivalently* $[\xi, \xi] = 0$*);*

iii) Jacobi identity: $[\xi\,, [\eta, \gamma]] - [\eta\,, [\xi, \gamma]] + [\gamma\,, [\xi, \eta]] = 0$ *;*

for $a, b \in \mathbb{R}$, $\xi, \eta, \gamma \in \mathfrak{g}$. *The **dual Lie algebra** $\mathfrak{g}^*$ is the dual vector space of $\mathfrak{g}$, and the natural pairing between $\mathfrak{g}$ and $\mathfrak{g}^*$ is*

$$\langle\,,\rangle : \mathfrak{g} \times \mathfrak{g}^* \to \mathbb{R}.$$

Remark 2.134. The dual Lie algebra $\mathfrak{g}^*$ is for us the dual vector space of $\mathfrak{g}$, that is the space of covectors that naturally pair with the elements in $\mathfrak{g}$. Much of the importance of $\mathfrak{g}^*$ for geometric mechanics comes from a natural Poisson structure that exists on $\mathfrak{g}^*$, and which can be connected to dynamics on the Lie group for a physical system with symmetry. Note that in the infinite dimensional case the choice of $\mathfrak{g}^*$ is no longer canonical, and the available freedom can lead to different equations of motions for the same physical system.

Remark 2.135. The Lie derivative $\pounds_X Y$ of a vector field was defined as $\pounds_X Y = [X, Y]$ for $X, Y \in \mathfrak{X}(\mathcal{M})$. The Jacobi identity is hence nothing but the Leibniz rule for the Lie derivative,

$$\pounds_X [Y, Z] = [\pounds_X Y, Z] + [Y, \pounds_X Z]$$

and it admits the usual interpretation of the product rule.

Example 2.87. The space of vector fields $\mathfrak{X}(\mathcal{M})$ on a manifold form a Lie algebra when equipped with the Jacboi-Lie bracket. Since $L_g^*[X_L, Y_L] = [L_g^* X_L, L_g^* Y_L] = [X_L, Y_L]$ for $X_L, Y_L \in \mathfrak{X}_L(G)$, the space $\mathfrak{X}_L(G)$ of left invariant vector fields forms a Lie subalgebra of $\mathfrak{X}(G)$, and analogously for right invariant vector fields.

When a Lie algebra arises from a group, the Lie bracket has to be chosen so that it is compatible with the group structure.

Definition 2.179. *The **Lie algebra** $\mathfrak{g}$ **of a Lie group** G is the tangent space T_eG at the identity together with the Lie bracket which is defined by **left extension***

$$[\xi, \eta] = [\xi, \eta]^L = [X_\xi, X_\eta]$$

by the Jacobi-Lie bracket $[\,,]$ of left invariant vector fields $X_\xi, X_\eta \in \mathfrak{X}_L(G)$ on G.

Remark 2.136. The Lie algebras defined by left and right extension coincide *up to a sign in the Lie bracket*, that is

$$[\xi, \eta]^R = -[\xi, \eta]^L.$$

This difference is important in mechanics for systems whose configuration space is a Lie group where the body (or convective) and spatial (or Eulerian) representations are obtained by left and right translation on the group from the Lagrangian representation.

Remark 2.137. For the Lie algebra of a Lie group, an alternative definition of the Lie bracket is possible by considering curves on G, much in the spirit of our original definition of a tangent space. Let $g(t)$ and $h(s)$ be curves on G satisfying $g(0) = e$ and $h(0) = e$ and $g'(0) = \xi$ and $h'(0) = \eta$. Then

$$[\xi, \eta] = \left. \frac{d}{dt}\frac{d}{ds}g(t)\,h(s)\,g^{-1}(t)\right|_{t=0,s=0}.$$

Intuitively, the Lie bracket hence measures the difference between the start and the endpoint for walking along the group first along $g^{-1}(t)$, then from the endpoint along $h(s)$, and then back using $g(t)$. Since the transformations on the group are only infinitesimal also the difference is only infinitesimal, and hence a vector. This also agrees with the anti-symmetry of the Lie-bracket, since we expect to end up exactly where we came from when we walk along one and the same curve.

Example 2.88. For a vector group V one has $T_e V \cong V$, and the Lie algebra $\mathfrak{v}$ coincides with the group. Since the invariant vector fields are the constant ones it follows that $[u, v] = 0$ and the Lie bracket vanishes for all $u, v \in \mathfrak{v}$. A Lie algebra for which the Lie bracket is trivial is sometimes denoted as an **Abelian Lie algebra**.

Next to the translation map and the left and right invariant vector fields that are obtained from it, the exponential map provides the principal tool that enables one to study problems in the Lie algebra and to transfer the result to the entire group, cf. Fig. 2.40.

Definition 2.180. *Let G be a Lie group and $\xi \in \mathfrak{g}$ be an element of the Lie algebra $\mathfrak{g}$ generating the left invariant vector field $X_\xi \in \mathfrak{X}_L(G)$. Denote by $\gamma_\xi : \mathbb{R} \to G$ the unique integral curve of X_ξ that satisfies $\gamma_\xi(0) = e$ and $\gamma'_\xi(t) = X_\xi(\gamma_\xi(t))$. The **exponential map** $\exp : \mathfrak{g} \to G$ **on** G is*

$$\exp(\xi) = \gamma_\xi(1)$$

and the maps generated by the left and right invariant vector fields coincide.
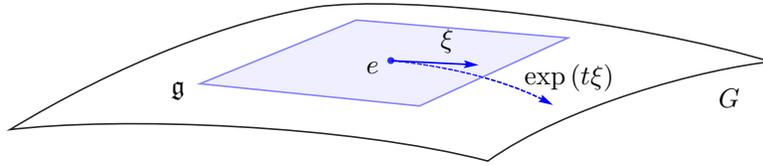
Figure 2.40: The exponential map $\exp : \mathfrak{g} \to G$ provides a mapping from the Lie algebra $\mathfrak{g}$ to the Lie group, and all one parameter subgroups of G are of the form $\exp t\xi$ for some $\xi \in \mathfrak{g}$.

That $\gamma_\xi(t)$ is defined for all $t \in \mathbb{R}$ follows from the group properties, and since the exponential map generated by the left and right invariant vector fields coincide, we will restrict us to the former ones in the following. The importance of the exponential map follows largely from the next proposition.

Proposition 2.80. *Let G be a Lie group with Lie algebra $\mathfrak{g}$. The integral curve $\gamma_\xi : \mathbb{R} \to G$ of the left invariant vector field $X_\xi \in \mathfrak{X}_L(G)$ is a one parameter subgroup of G*

$$\gamma(s + t) = \gamma(s)\,\gamma(t)$$

and, moreover, $\exp(t\xi) = \gamma_\xi(t)$ is t-linear. Hence, the exponential map provides a homomorphism between the one parameter subgroups $t\xi : \mathbb{R} \to \mathfrak{g}$ of $\mathfrak{g}$ and $\gamma_\xi(t) : \mathbb{R} \to G$ of G.

The above proposition enables to consider linear vector addition in the Lie algebra $\mathfrak{g}$ instead of the nonlinear group multiplication on the Lie group G.

Remark 2.138. On a Riemannian manifold, the exponential map is defined using geodesics. On a Lie group admitting a bi-invariant metric, that is a metric that is invariant under the left and right translation action, both notions coincide.[128] The one parameters subgroups $\gamma_\xi(t)$ can therefore be considered as generalized geodesics on a Lie group.

The following proposition shows that one parameter subgroups are in one-to-one correspondence with invariant vector fields.

Proposition 2.81. *All one parameter subgroups of a Lie group G are generated by an invariant vector field $X_\xi \in \mathfrak{X}_L(G)$ and of the form $\exp(t\xi)$.*

Since the exponential map is smooth, which follows from the smoothness of the group operations, and since $T_0 \exp(0) = \mathrm{id}_\mathfrak{g}$ is an isomorphism of vector spaces, the inverse function theorem implies the following result.

Proposition 2.82. *The exponential map provides a local diffeomorphism from a neighborhood of zero in $\mathfrak{g}$ to a neighborhood of the identity in G.*

---

[128]See for example (Montgomery, *A Tour of Subriemannian Geometries, Their Geodesics, and Applications*) for details.

The exponential map can hence be used to define a chart around the identity, and using the translation action this chart can be used to generate an atlas for G, cf. Remark 2.133.

Example 2.89. For a vector group V the exponential map $\exp : \mathfrak{v} \cong V \to V : v \to v$ is the identity. The one parameter subgroups generated by the exponential map have the form

$$\exp(tv) = tv$$

of rays centered at the origin.

As the following result shows, the Lie algebra and the exponential maps are naturally operations for a Lie group.

Proposition 2.83. *Let G and H be Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively, and let $\varphi : G \to H$ be a Lie group homomorphism so that $\varphi(g_1 \circ g_2) = \varphi(g_1) \circ \varphi(g_2)$ for $g_1, g_2 \in G$. Then*

$$T_e\varphi : T_e G \to T_e H := \mathfrak{g} \to \mathfrak{h}$$

is a Lie algebra homomorphism so that $(T_e\varphi)([\xi, \eta]) = [(T_e\varphi)\xi, (T_e\varphi)\eta]$ for $\xi, \eta \in g$, and

$$\varphi(\exp(t\xi)) = \exp\left((T_e\varphi)\xi\right).$$

Example 2.90. The determinant $\det : L(\mathbb{R}^n, \mathbb{R}^n) \to \mathbb{R} : A \to \det(A)$ defines a group homomorphism between $\mathrm{GL}(n, \mathbb{R})$ and the multiplicative group $(\mathbb{R}, \cdot)$ since

$$\det(A\,B) = \det(A)\det(B)$$

for $A, B \in \mathrm{GL}(n, \mathbb{R})$. Proposition 2.83 applied to this map yields

$$\det\left(\exp(A)\right) = \exp\left(\mathrm{tr}(A)\right).$$

where the trace $\mathrm{tr} : \mathfrak{gl}(n) \to \mathbb{R}$ is the Lie algebra homomorphism from the Lie algebra $\mathfrak{gl}(n)$ of $\mathrm{GL}(n, \mathbb{R})$ to the Lie algebra $(\mathbb{R}, +)$ of $(\mathbb{R}, \cdot)$ associated with the determinant map. The example also shows that a Lie algebra homomorphism does no imply a vector space isomorphism.

Another concept related to group homomorphisms that will become important in Sec. 2.3.3.3 when we consider infinite dimensional Lie groups is the following.

Definition 2.181. *Let $\{G_i\}_{i=0}^{\infty}$ be a sequence of groups that are related by surjective homomorphisms $f_i : G_i \to G_{i-1}$. The **inverse limit***

$$\varprojlim G_i = \left\{ \{(g_0, g_1, \ldots)\} \mid f_i(g_i) = g_{i-1},\ g_i \in G_i, \{(g_0, g_1, \ldots)\} \in \prod_{i=0}^{\infty} G_i \right\}$$

is the subset $\{(g_0, g_1, \ldots)\}$ of the tensor product $\prod_{i=0}^{\infty} G_i$ satisfying $f_i(g_i) = g_{i-1}$.
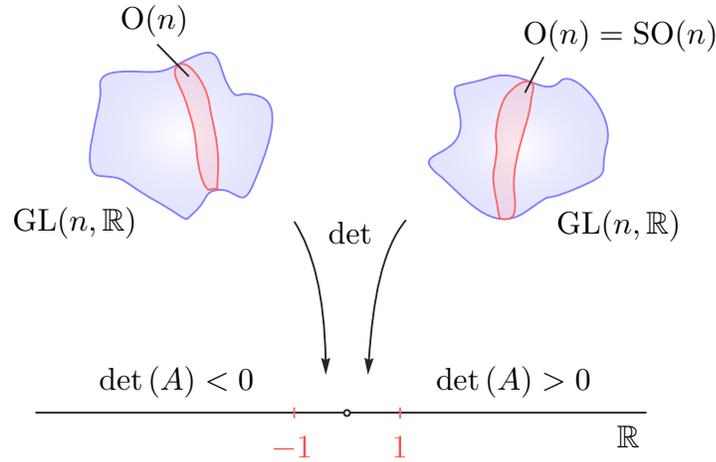
Figure 2.41: The determinant $\det : \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}$ provides a group homomorphism from the general linear group $\mathrm{GL}(n, \mathbb{R})$ to $(\mathbb{R}, \cdot)$.

Intuitively, the structure provided by an inverse limit always allows to "go up one level" from G_i to G_{i+1} if necessary and use the additional structure available on G_{i+1}. The next definition uses the concept of an injectively immersed submanifold that was introduced in Def. 2.81.

Definition 2.182. *Let G be a Lie group. A **Lie subgroup** H of G is a subgroup under multiplication in G and an injectively immersed submanifold of G. A subgroup H that is also a submanifold for G is called **regular**.*

The distinction between a subgroup and a Lie subgroup is important in many context. The Lie algebra of a Lie subgroup is characterized in the next proposition.

Proposition 2.84. *Let G be a Lie group with Lie algebra $\mathfrak{g}$, and H a Lie subgroup of G with Lie algebra $\mathfrak{h}$. Then $\mathfrak{h}$ is a Lie subalgebra of $\mathfrak{g}$ and*

$$\mathfrak{h} = \{\xi \in \mathfrak{g} \mid \exp(t\xi) \in H \text{ for all } t \}.$$

Example 2.91. Let us consider again the general linear group $\mathrm{GL}(n, \mathbb{R})$, the group of invertible maps in $L(\mathbb{R}^n, \mathbb{R}^n)$, and continue Example 2.85. It can be shown that the group is noncompact, and from the determinant map $\det : \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}$, which provides a group homomorphism from $\mathrm{GL}(n, \mathbb{R})$ to $\mathbb{R}$, cf. Example 2.90, it follows that $\mathrm{GL}(n, \mathbb{R})$ has two connected components corresponding to $\det(A) < 0$ and $\det(A) > 0$, cf. Fig. 2.41. The Lie algebra $\mathfrak{gl}(n)$ of $\mathrm{GL}(n, \mathbb{R})$ is given by

$$\mathfrak{gl}(n) = (Z \in L(\mathbb{R}^n, \mathbb{R}^n), [Z, Y] = Z\,Y - Y\,Z),$$

and it is the set of all $n \times n$ matrices with the Lie bracket given by the usual commutator for matrices. The exponential map for $\mathrm{GL}(n,\mathbb{R})$ is the matrix expoential

$$e^Z = \sum_{i=0}^{\infty} \frac{Z^i}{i!} = I + Z + \frac{1}{2}Z^2 + \dots$$

where $Z \in \mathfrak{gl}(n)$, and $e^{tZ} : \mathbb{R} \to \mathrm{GL}(n,\mathbb{R})$ forms a one parameter subgroup. Various Lie subgroups of the general linear group are of interest. The **special linear group**

$$\mathrm{SL}(n) = \{A \in \mathrm{GL}(n,\mathbb{R}) \mid \det(A) = 1\} \subset \mathrm{GL}(n,\mathbb{R})$$

represents all volume preserving transformations on $\mathbb{R}^n$. Its Lie algebra is

$$\mathfrak{sl}(n) = \{Z \in \mathfrak{gl}(n) \mid \mathrm{tr}(Z) = 0\}$$

and it inherits the Lie bracket from $\mathrm{GL}(n,\mathbb{R})$ since it is a Lie subgroup. All orthogonal transformations of Euclidean space form the **orthogonal group**

$$\mathrm{O}(n) = \{A \in \mathrm{GL}(n,\mathbb{R}) \mid \langle Ax, Ay \rangle = \langle x,y \rangle\} \subset \mathrm{GL}(n,\mathbb{R})$$

where $\langle\,,\rangle$ denotes the usual dot product in $\mathbb{R}^n$. The Lie algebra $\mathfrak{o}(n)$ of $\mathrm{O}(n)$ is given by

$$\mathfrak{o}(n) = \left\{Z \in \mathfrak{gl}(n) \mid Z = -Z^T\right\},$$

and it is formed by the anti-symmetric matrices, which are necessarily also traceless. The **special orthogonal group** $\mathrm{SO}(n)$, which represents all rotations of $\mathbb{R}^n$ or all orientation preserving orthgonal transformations, is

$$\mathrm{SO}(n) = \mathrm{SL}(n) \cap \mathrm{O}(n)$$

and it is the connected component of $\mathrm{O}(n)$ containing the identity. The Lie algebra $\mathfrak{so}(n)$ of the special orthogonal group $\mathrm{SO}(n)$ hence conincides with those of $\mathrm{O}(n)$, that is

$$\mathfrak{so}(n) = \left\{Z \in \mathfrak{gl}(n) \mid Z = -Z^T\right\}.$$

In the following, we will also need some elementary notions of quotients of Lie groups, and these are introduced next.

Definition 2.183. *Let H be a closed subgroup of the Lie group G. The **Lie group quotient** G/H is the collection of all cosets $[g]$ of G under H, that is*

$$G/H = \{[g] \in G \mid g\,H \in [g]\},$$

with quotient projection $\pi : G \to G/H = g \to [g]$.

The group quotient hence identifies all elements in G that differ by an element in H, and the "curve" in G traced out by gH is *one* element of the quotient G/H. The next theorem justifies the use of the term *Lie* group quotient.

Theorem 2.28. *Let H be a closed subgroup of the Lie group G. Then G/H has a unique manifold structure such that $\pi : G \to G/H$ is a smooth surjective submersion.*

Example 2.92. The quotient $SO(3)/SO(2)$ identifies all rotations around a fixed axis. The quotient $SO(3)/SO(2)$ is hence the space of all rotation axes, which is isomorphic to the 2-sphere S^2 and has a smooth manifold structure.

Remark 2.139. Every locally compact group admits, up to a multiplicative constant, a (left) translation invariant measure known as **Haar measure**. For example, for the vector group $(\mathbb{R}^n, +)$ the Haar measure is the usual Lebesgue measure. For any Lie group G, the Haar measure can be constructed using the translation on the group: Let $\omega \in \bigwedge^n(T_e G)$ be a volume form in the tangent space $T_e G$ at the identity. Then $\omega = L_g^* \omega$ defines a non-degenerate volume form on all of G. A corollary of this construction is that every Lie group is orientable.

2.3.3.2 Actions of Lie Groups

In applications, Lie groups arise when these act on spaces,[129] and for the study of dynamics on Lie group also the actions of a Lie group on itself and its Lie algebra are important. Actions of Lie groups on manifolds and vector spaces will be studied in the following.

Definition 2.184. *The **left action of the Lie group** G **on a manifold** $\mathcal{M}$ is a smooth mapping $\varphi : G \times \mathcal{M} \to \mathcal{M}$ satisfying*

i) identity: $\varphi(e, x) = x$, for all $x \in \mathcal{M}$;

ii) composition: $\varphi(g, \varphi(h, x)) = \varphi(gh, x)$, for all $g, h \in G$ and $x \in \mathcal{M}$.

Remark 2.140. A right action of a Lie group G on a manifold $\mathcal{M}$ is defined analogously with the composition satisfying $\varphi(\varphi(x, g), h) = \varphi(x, gh)$.

The action of a Lie group on a manifold is usually abbreviated as $\varphi(g, x) = g \circ x = gx$ for a left action, and $\varphi(x, g) = x \circ g = xg$ for a right action.

---

[129] Arnold (*On Teaching Mathematics*) emphasized the importance of the action of groups on spaces as follows: "What is a group? Algebraists teach that this is supposedly a set with two operations that satisfy a load of easily-forgettable axioms. This definition provokes a natural protest: why would any sensible person need such pairs of operations? 'Oh, curse this maths' - concludes the student (who, possibly, becomes the Minister for Science in the future). We get a totally different situation if we start off not with the group but with the concept of a transformation (a one-to-one mapping of a set onto itself) as it was historically. A collection of transformations of a set is called a group if along with any two transformations it contains the result of their consecutive application and an inverse transformation along with every transformation."

Remark 2.141. Much of the previous discussion on Lie groups can be considered from the point of view of group actions. For example, the circle group in Example 2.83 can be considered as the action of $\mathbb{R}$ on S^1.

Remark 2.142. Considered as a map $\varphi_g : \mathcal{M} \to \mathcal{M}$ for a fixed element $g \in G$, the action of G on $\mathcal{M}$ satisfies $\varphi_e = \mathrm{id}(\mathcal{M})$ and $\varphi_{gh} = \varphi_g \circ \varphi_h$. **An action is hence a group homomorphism from G into the diffeomorphisms $\mathrm{Diff}(\mathcal{M})$ of $\mathcal{M}$.** The action of G on a manifold can be generalized to the action on other objects by replacing diffeomorphisms by isomorphisms in the appropriate category.

When a Lie group acts on a vector space the action is known as a representation.[130] Such actions are studied in representation theory, which is an important part of the general theory of Lie groups.

Definition 2.185. *A **representation** of a Lie group G is the action of G on a Banach space $\mathbf{E}$ by continuous linear transformations $\varphi_g : \mathbf{E} \to \mathbf{E}$.*

Example 2.93. Rotation matrices $\mathrm{R} : \mathbb{R}^3 \to \mathbb{R}^3$ are a representation of the rotation group $\mathrm{SO}(3)$ on $\mathbb{R}^3$, cf. Example 2.91. The representation of the Lie algebra $\mathfrak{so}(3)$ are the skew-adjoint matrices, and these can be identified with a vector in $\mathbb{R}^3$ by a vector space isomorphism known as **hat map**

$$\mathbb{R}^3 \to \mathfrak{so}(3) : \mathrm{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \to \hat{\mathrm{v}} = \begin{pmatrix} 0 & -v_1 & v_2 \\ v_1 & 0 & -v_3 \\ -v_2 & v_3 & 0 \end{pmatrix}. \qquad (2.189)$$

Using the hat map, the Lie bracket is given by the vector cross product, that is $[\hat{\mathrm{u}}, \hat{\mathrm{v}}] = \mathrm{u} \times \mathrm{v}$ for $\hat{\mathrm{u}}, \hat{\mathrm{v}} \in \mathfrak{so}(3)$. It hence follows that $\hat{\cdot} : \mathbb{R}^3 \to \mathfrak{so}(3)$ also provides a Lie algebra homomorphism. The inverse of the hat map is the **check map** and it is given by

$$\mathfrak{so}(3) \to \mathbb{R}^3 : \mathrm{v} = \begin{pmatrix} 0 & -v_1 & v_2 \\ v_1 & 0 & -v_3 \\ -v_2 & v_3 & 0 \end{pmatrix} \to \check{\mathrm{v}} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \qquad (2.190)$$

and it is easy to see that the hat and the check map are in fact inverse operations.

Example 2.94. Let $\mathcal{M}$ be a manifold and $X \in \mathfrak{X}(\mathcal{M})$ be a complete vector field on $\mathcal{M}$. Then the flow F_t associated with X defines an action $F_t : \mathbb{R} \times \mathcal{M} \to \mathcal{M}$ of $\mathbb{R}$ on $\mathcal{M}$.

Another important special case of Lie group actions are semi-direct products.

Definition 2.186. *Let G be a Lie group and V a vector space with the usual group structure under addition. The **semi-direct product** $G \ltimes V$ is a Lie group with multiplication*

$$(g_2, v_2) \cdot (g_1, v_1) = (g_2 g_1, g_2 v_1 + v_2).$$

---

[130]We will restrict us to Banach spaces since otherwise, for example in a Fréchet space, the notion of differentiability, necessary to characterize smoothness, is not well defined.

for $g_1, g_2 \in G$ and $v_1, v_2 \in V$. The inversion map for a semi-direct product is defined by $\text{Inv}(g, v) = (g^{-1}, -g^{-1}v)$ and the identity is $(e, 0)$.

Example 2.95. Let $SO(3) \ltimes \mathbb{R}^3$ be the semi-direct product of the rotation group $SO(3)$ and Euclidean space $\mathbb{R}^3$. The group action is then

$$(R_2, x_2) \cdot (R_1, x_1) = (R_2 R_1, R_2 x_1 + x_2)$$

for $R_1, R_2 \in SO(3)$ and $x_1, x_2 \in \mathbb{R}^3$. $SO(3) \ltimes \mathbb{R}^3$ is hence the Lie group of affine transformations.

The following definition summarizes important properties of Lie group actions.

Definition 2.187. *Let G be a Lie group acting on the manifold $\mathcal{M}$. The **orbit** $\text{Orb}_G(\bar{x}) \subset \mathcal{M}$ **of** $\bar{x} \in \mathcal{M}$ **under** G*

$$\text{Orb}_G(\bar{x}) = \{x \in \mathcal{M} \mid x = g \circ \bar{x}, g \in G\}$$

*is the set of all elements in $\mathcal{M}$ that are accessible from $\bar{x}$ under G. The **stabilizer** (or isotropy or symmetry) $G_{\bar{x}} \subset G$ of the group G at $\bar{x} \in \mathcal{M}$ is the Lie subgroup*

$$G_{\bar{x}} = \{g \in G \mid g \circ \bar{x} = \bar{x}\}$$

that leaves $\bar{x}$ invariant. Moreover, an action $\varphi : G \times \mathcal{M} \to \mathcal{M}$ is

  i) ***transitive***, *if there is only one orbit and for every $x, y \in \mathcal{M}$ there exists a $g \in G$ such that $y = g \circ x$;*

  ii) ***effective*** *(or faithful), if $\varphi_g = \text{id}(\mathcal{M})$ implies $g = e$ and $g \to \varphi_g$ is one-to-one;*

  iii) ***free***, *if $\varphi_g(x) = x$ implies $g = e$ for all $x \in \mathcal{M}$; equivalently, an action is free if and only if the stabilizer $G_x = \{e\}$ for all $x \in \mathcal{M}$, and when an action is free then it is also effective;*

  iv) ***proper***, *if for every convergent sequence $\{x_n\}$ in $\mathcal{M}$ for which $\varphi_{g_n}(x_n)$ converges then $\{g_n\}$ has a convergent subsequence in G, in particular the action of any compact group is always proper.*

Remark 2.143. Effective and free actions can also be characterized as follows. An action is

  i) *effective, if $\nexists g \in G, g \neq e : \varphi_g = \text{id}(\mathcal{M})$;*

  ii) *free, if $\nexists x \in \mathcal{M}, g \neq e : \varphi_g \circ x = x$.*

The notions correspond to non-degenerate actions and lead to "well behaved" symmetries in physical systems.

Example 2.96. Consider the action of $\mathrm{SO}(3)$ on the 2-sphere S^2. The stabilizer $\mathrm{SO}_{\bar{\omega}}(3)$ for a point $\bar{\omega} \in S^2$ is the set of all rotations in $\mathrm{SO}(3)$ that leave $\bar{\omega}$ invariant, that is the set of rotations whose axis goes through $\bar{\omega}$. Since the stabilizer $\mathrm{SO}_{\bar{\omega}}(3)$ is non-trivial, the action of $\mathrm{SO}(3)$ on S^2 is not free, although it is transitive by Euler's theorem.

Example 2.97. The left translation action $L_g : G \times G \to G$ of a Lie group on itself is a transitive action that is free, and hence also faithful. Right translation $R_g : G \times G \to G$ defines a right action of a Lie group on itself, and $R_{g^{-1}}$ defines again a left action.

As the above example demonstrates, some of the most important actions of Lie groups are when the group is acting on itself or its Lie algebra. The next definition considers the latter case.

Definition 2.188. *The **adjoint action** $\mathrm{Ad} : G \times \mathfrak{g} \to \mathfrak{g}$ of a Lie group G on its Lie algebra $\mathfrak{g}$ is*

$$\mathrm{Ad}_g = T_e \mathrm{Inv}_g : \mathrm{Ad}_g(\xi) = \left(T_g R_{g^{-1}} \circ T_e L_g \right) \xi.$$

The adjoint action Ad can hence be seen as the infinitesimal version of the inversion map.

Definition 2.189. *The **coadjoint action** $\mathrm{Ad}^* : G \times \mathfrak{g}^* \to \mathfrak{g}^*$ of a Lie group G on is dual Lie algebra $\mathfrak{g}^*$ is*

$$\langle \mathrm{Ad}(\xi), \alpha \rangle = \langle \xi, \mathrm{Ad}^*(\alpha) \rangle$$

for $\xi \in \mathfrak{g}$ and $\alpha \in \mathfrak{g}^*$, and where $\langle , \rangle$ denotes the natural pairing between elements in $\mathfrak{g}$ and $\mathfrak{g}^*$.

The coadjoint action can hence be considered as the adjoint of the Ad-action.

Remark 2.144. Analogous to the coadjoint action in Def. 2.189, every representation of a Lie group G on a vector space V induces a *contragredient* representation on the dual space V^*. The coadjoint action Ad^* is hence a special case of a contragredient action.

The orbit $\mathrm{Orb}_G(x)$ of a point $x \in \mathcal{M}$ defines an equivalence class $x \sim y$ and the quotient space $\mathcal{M}/G$ of all orbits reduces $\mathcal{M}$ by the symmetries described by G. We formalize these notions in the following definition.

Definition 2.190. *Let G be a Lie group acting on a manifold $\mathcal{M}$. The **orbit space***

$$\mathcal{M}/G = \{ x \sim y \mid y \in \mathrm{Orb}_G(x) \}$$

is the set of all equivalence classes generated by the action of G on $\mathcal{M}$, and it has the natural projection $\pi : \mathcal{M} \to \mathcal{M}/G$.

Unfortunately, the orbit space is well behaved only when certain conditions are satisfied, as the next proposition shows, cf. Def. 2.187.

Proposition 2.85. *Let G be a Lie group acting on a manifold $\mathcal{M}$ by $\varphi : G \times \mathcal{M} \to \mathcal{M}$. If φ is proper and free then the orbit space $\mathcal{M}/G$ is a smooth manifold and $\pi : \mathcal{M} \to \mathcal{M}/G$ is a smooth submersion.*

Example 2.98. With Euler's theorem in mind, consider a rotation around the z-axis; that is the action of the $\mathrm{SO}_z(3)$ on S^2. Using spherical coordinate, the orbit space is then naturally identified with $[0, \pi]$ since all points on the same latitude on S^2 are in the same orbit under $\mathrm{SO}_z(3)$.

Example 2.99. Let $G = \mathbb{R}$ act on $\mathcal{M} = \mathbb{R}$ by translation so that $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R} : (u, v) = u + v$. The action is transitive, free, and proper, and since $\mathrm{Orb}_\mathbb{R}(\bar{x}) = \mathbb{R}$ the orbit space $\mathcal{M}/G = \mathbb{R}/\mathbb{R}$ is a single point.

Infinitesimally, the action of a Lie group is characterized by the infinitesimal generator, which is introduced next.

Definition 2.191. *Let G be a Lie group that acts by $\varphi : G \times \mathcal{M} \to \mathcal{M}$ on a manifold $\mathcal{M}$, and let*

$$\varphi^\xi(t, x) = \exp(t\xi) \circ x$$

be the $\mathbb{R}$-action generated by $\xi \in \mathfrak{g}$. Then $\varphi_t^\xi : \mathbb{R} \times \mathcal{M} \to \mathcal{M}$ is a flow on $\mathcal{M}$ and the corresponding vector field on $\mathcal{M}$

$$\xi_\mathcal{M}(x) = \frac{d}{dt} \left(\exp\left(t\xi \right) \circ x \right) \bigg|_{t=0} \in \mathfrak{X}(\mathcal{M})$$

*is the **infinitesimal generator of** φ_t^ξ.*

The infinitesimal generator $\xi_\mathcal{M} : \mathfrak{g} \to T\mathcal{M}$ can hence be considered as the infinitesimal action of G on $\mathcal{M}$, or as the action of the Lie algebra ξ on $\mathcal{M}$, and this is made more precise in the following proposition.

Proposition 2.86. *The exponential map provides a Lie algebra anti-homomorphism from $\mathfrak{g}$ to a Lie subalgebra of $\mathfrak{X}(\mathcal{M})$ with bracket*

$$[\xi, \eta]_\mathcal{M} = -[\xi_\mathcal{M}, \eta_\mathcal{M}].$$

With the above results it is natural to define the action of a Lie algebra on a manifold as follows.

Definition 2.192. *Let $\mathfrak{g}$ be a Lie algebra and $\mathcal{M}$ a manifold. The **Lie algebra action** $\phi : \mathfrak{g} \times \mathcal{M} \to T\mathcal{M}$ is a Lie algebra anti-homomorphism from $\mathfrak{g}$ into $(\mathfrak{X}(\mathcal{M}), [,])$ such that*

$$\phi(\xi, x) \to \xi_\mathcal{M}(x)$$

is smooth for $\xi \in \mathfrak{g}$ and $x \in \mathcal{M}$.

Lie group actions can always be formulated using their infinitesimal analogoue, and the vector field description is often more insightful, similar to the situation for arbitrary flows in the foregoing that can also either be given as a vector field or as a time evolution map. Hence, in many applications it is natural and useful to employ Lie algebra actions instead of Lie group actions.

Remark 2.145. There is a one-to-one correspondence for representations of Lie groups and Lie algebras, that is given the representation of a Lie group there is a unique representation of the Lie algebra associated with the group.

The following proposition characterizes the subspace of $\mathfrak{X}(\mathcal{M})$ spanned by infinitesimal generators associated with a group action.

Proposition 2.87. *Let G be a Lie group acting on a manifold $\mathcal{M}$, and let $\bar{x}$ be a fixed element in $\mathcal{M}$. The tangent space $T_x\mathrm{Orb}(\bar{x})$ of the orbit $\mathrm{Orb}(\bar{x})$ at $x \in \mathrm{Orb}(\bar{x})$ is*

$$T_x\mathrm{Orb}(\bar{x}) = \{\xi_{\mathcal{M}} \mid \xi \in \mathfrak{g}\}.$$

Moreover, the map $G/G_{\bar{x}} \to \mathrm{Orb}(\bar{x})$ from the quotient $G/G_{\bar{x}}$ of G and the stabilizer $G_{\bar{x}}$ to the orbit $\mathrm{Orb}(\bar{x})$ is a diffeomorphism.

Example 2.100. Consider again Example 2.96. The stabilizer $\mathrm{SO}_{\bar{\omega}}(3)$ are the rotation whose axis is $\bar{\omega} \in S^2$. Hence in $\mathrm{SO}(3)/\mathrm{SO}_{\bar{\omega}}(3)$ two rotations are identified when they differ by a rotation around $\bar{\omega}$. But this are exactly the rotations that have a one-to-one correspondence to points $\omega \in S^2$ which is the orbit of $\bar{\omega}$ under $\mathrm{SO}(3)$.

Corollary 2.17. *The Lie algebra $\mathfrak{g}_x$ of the stabilizer group G_x at $x \in \mathcal{M}$ is*

$$\mathfrak{g}_x = \{\xi \in \mathfrak{g} \mid \xi_{\mathcal{M}}(x) = 0\}.$$

Example 2.101. The Lie algebra $\mathfrak{s} = \mathfrak{g} \ltimes V$ of a semi-direct product $G \ltimes V$, cf. Def. 2.186, has the bracket

$$[(\xi_1, v_1), (\xi_2, v_2)] = ([\xi_1, \xi_2], \xi_1 v_2 - \xi_2 v_1)$$

for $\xi_1, \xi_2 \in \mathfrak{g}$ and $v_1, v_2 \in V$, where ξv denotes the infinitesimal generator of ξ on V.

The most important examples for infinitesimal generators arise again from the action of a Lie group on its Lie algebra.

Definition 2.193. *Let $\mathrm{Ad} : G \times \mathfrak{g} \to \mathfrak{g}$ be the adjoint action of a Lie group G on its Lie algebra $\mathfrak{g}$. The **ad operator** $\mathrm{ad} : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ is the infinitesimal generator $\xi_{\mathfrak{g}}$ of Ad given by*

$$\mathrm{ad}_\xi(\eta) = \xi_{\mathfrak{g}}(\eta) = [\xi, \eta].$$

With Remark 2.137 and the definition of the adjoint action it comes at no surprise that the infinitesimal generator $\mathrm{ad}_\xi(\eta)$ of $\mathrm{Ad}_{\exp(t\xi)}(\eta)$ is given by the Lie bracket, and this also provides additional intuition for $[\,,]$. For the coadjoint action a similar result holds.

Definition 2.194. *Let* $\mathrm{Ad}^* : G \times \mathfrak{g}^* \to \mathfrak{g}^*$ *be the coadjoint action of a Lie group* G *on its dual Lie algebra* $\mathfrak{g}^*$. *The* ***ad**^* **operator*** $\mathrm{ad}^* : \mathfrak{g} \times \mathfrak{g}^* \to \mathfrak{g}^*$ *is the infinitesimal generator of the coadjoint action* Ad^* *given by*

$$\langle \xi_{\mathfrak{g}}(\eta), \alpha \rangle = \langle \mathrm{ad}_\xi(\eta), \alpha \rangle = \langle \eta, \mathrm{ad}_\xi^* \alpha \rangle .$$

for arbitrary $\alpha \in \mathfrak{g}^*$.

The ad^* operator is central to the description of dynamics in the dual Lie algebra for Hamiltonian systems with symmetry. Important in this context is also the following definition which establishes when a map between manifolds is natural with respect to a group action.

Definition 2.195. *Let* G *be a Lie group that acts on manifolds* $\mathcal{M}$ *and* $\mathcal{N}$ *by* $\varphi : G \times \mathcal{M} \to \mathcal{M}$ *and* $\psi : G \times \mathcal{N} \to \mathcal{N}$. *A smooth map* $f : \mathcal{M} \to \mathcal{N}$ *is* ***equivariant*** *if*

$$f(\varphi \circ x) = \psi \circ f(x).$$

A mapping is hence equivariant when it commutes with the group action, and the following proposition shows that this also carries over to the Lie algebra.

Proposition 2.88. *Let* G *be a Lie group and* $f : \mathcal{M} \to \mathcal{N}$ *be an equivariant map between manifolds on which* G *acts. Then the infinitesimal generators* $\xi_{\mathcal{M}}$ *and* $\xi_{\mathcal{N}}$ *are* f-*related so that* $Tf \circ \xi_{\mathcal{M}} = \xi_{\mathcal{N}} \circ f$, *and if* f *is a diffeomorphism then* $f^* \xi_{\mathcal{N}} = \xi_{\mathcal{M}}$.

An important notion of an induced action of a Lie group on a manifold $\mathcal{M}$ is the following, and it is a special case of a cotangent lift which was introduced in Def. 2.95.

Definition 2.196. *Let* $\varphi : G \times \mathcal{M} \to \mathcal{M}$ *be the action of a Lie group* G *on the manifold* $\mathcal{M}$. *The* ***right lifted action*** $\varphi^* : G \times T^*\mathcal{M} \to T^*\mathcal{M}$ *is the right action of* G *on the cotangent bundle* $T^*\mathcal{M}$ *defined by*

$$\varphi_g^*(\alpha) = (T_{g^{-1}m}^* \varphi_g)(\alpha)$$

where $g \in G$ *and* $\alpha \in T_m^*\mathcal{M}$. *The* ***left lifted action*** $\varphi_* : G \times T^*\mathcal{M} \to T^*\mathcal{M}$ *is defined by*

$$\varphi_*^g(\alpha) = (T_{gm}^* \varphi_{g^{-1}})(\alpha)$$

where $\alpha \in T_{g^{-1}m}^*\mathcal{M}$.

That φ^* is indeed a right action follows from the properties of the pullback, cf. Proposition 2.48.

Example 2.102. Let $\mathcal{M} = \mathbb{R}^3$ and let the vector Lie group $(\mathbb{R}^3, +)$ act by $z \circ x \mapsto x + z$ for $x \in \mathbb{R}^3$ and $z \in (\mathbb{R}^3, +)$. The lifted action of $(\mathbb{R}, +)$ on the cotangent bundle $T^* \mathbb{R}^3$ is then $z \circ (x, p) \mapsto (x + z, p)$ and the fiber variable p is unaffected by the action.

Example 2.103. Let $\mathcal{M} = \mathbb{R}^3$ and let the rotation group $\mathrm{SO}(3)$ act on the space as usual. The lifted action of $\mathrm{SO}(3)$ on $T^* \mathbb{R}^3$ is then $R \circ (x, p) = (R\,x, R\,p)$ for $(x, p) \in T^* \mathbb{R}^3$ and $R \in \mathrm{SO}(3)$.

2.3.3.3 Infinite Dimensional Lie Groups[131]

Infinite dimensional Lie groups arise in many applications in particular in continuum mechanics, although the theory of such groups is currently far from satisfactory. Following the literature, we will treat infinite dimensional Lie groups formally, assuming all results from the finite dimensional case hold. Some of the technical aspects will, however, be discussed in this section.

Concept and Conceptual Difficulties An infinite dimensional Lie group is an infinite dimensional manifold that is equipped with smooth group multiplication and inversion. One of the principal reasons for the perhaps unusual general development of the manifold concept in Sec. 2.3.2, using Banach spaces as model spaces, was to provide a foundation for infinite dimensional Lie groups. In the case of a Banach or Hilbert manifold, that is when the model space is a Banach or Hilbert space, many of the results of the finite dimensional theory still hold, while for Fréchet manifolds, which are sometimes considered in the mathematics literature,[132] neither the inverse nor the implicit function theorem are available, and not even a canonical notion of differentiability exists. However, even in the case of Banach or Hilbert manifolds, the smoothness of the group operations is by no means guaranteed, and there are important examples where it does not hold. Additional conveniences of the finite theory that are in general no longer satisfied in the infinite dimensional setting are:

- the exponential map is not a local diffeomorphism onto a neighborhood of the identity;

---

[131]For some up-to-date surveys, which also consider the theory from different vantage points, see (Neeb, "Towards a Lie theory of locally convex groups"; Smolentsev, "Diffeomorphism groups of compact manifolds"; Eichhorn, "Partial differential equations on closed and open manifolds"; Schmid, "Infinite Dimensional Lie Groups and Algebras in Mathematical Physics"), the last reference also gives an introduction into the use of (infinite-dimensional) Lie groups into "modern" physics, including super-symmetry. Some of the classical references on the subject are (Ebin and Marsden, "Groups of Diffeomorphisms and the Motion of an Incompressible Fluid"; Chernoff and Marsden, "On continuity and smoothness of group actions"; Chernoff and Marsden, *Properties of Infinite Dimensional Hamiltonian Systems*; Ratiu and Schmid, "The differentiable structure of three remarkable diffeomorphism groups"; Marsden and Hughes, *Mathematical Foundations of Elasticity*; Adams, Ratiu, and Schmid, "A Lie group structure for pseudodifferential operators"; Adams, Ratiu, and Schmid, *A Lie group structure for Fourier integral operators*).

[132]See for example (Khesin and Wendt, *The Geometry of Infinite-Dimensional Groups*).

- a closed subgroup of G is not necessarily a Lie subgroup of G, that is it does not necessarily have a manifold structure;

- not every infinite dimensional Lie algebra has (at least) one associated Lie group whose infinitesimal generator it is;

- the dual Lie algebra is no longer uniquely defined.

The last point is particular important for physical systems whose configuration space is an infinite dimensional Lie group, and it can lead to different equations for the dynamics depending on the choice of the dual Lie algebra; an example is ideal fluid dynamics and we will discuss it in detail in Example 2.147.

The most important examples of infinite-dimensional Lie groups for our purposes are diffeomorphism groups. However, many other infinite dimensional groups are of importance in mathematical physics as the following examples show.

Example 2.104. Invertible, bounded linear operators in $L(V, V)$ on an infinite dimensional Banach space V form an infinite dimensional Lie group, and it provides the infinite dimensional analogue of the general linear group $\mathrm{GL}(n, \mathbb{R})$ on $\mathbb{R}^n$ that was discussed in Example 2.91. An analogue of $\mathrm{SO}(n)$ is hence the group $\mathrm{U}(\mathcal{H})$ formed by unitary operators on a Hilbert space $\mathcal{H}$, which was studied extensively starting in the 1930s for applications in quantum mechanics. Often one thereby considers a subgroup that arises from a group homomorphism $\rho : G \to \mathrm{U}(\mathcal{H})$, and this is known as a unitary representation, since G is acting on the vector space $\mathcal{H}$. For example, using the group theoretic perspective of this section, the one parameter groups U_t of unitary operators that were introduced in Sec. 2.3.2.7 can be interpreted as the action $\mathbb{R} \times \mathcal{H} \to \mathcal{H}$. Stone's theorem, which in the complex case asserts a one-to-one correspondence between one parameter groups of unitary operators and self-adjoint operators, cf. Remark 2.127, then shows that the Lie algebra $\mathfrak{u}(\mathcal{H})$ of $\mathrm{U}(\mathcal{H})$ can be identified with the space of self-adjoint operators. Moreover, we also see that the exponential $U_t = e^{tA}$ that defines the one parameter subgroup U_t is the usual exponential map for the group $\mathrm{U}(\mathcal{H})$. Hence, one can consider Stone's theorem also as a consequence of the infinite dimensional Lie group structure of $\mathrm{U}(\mathcal{H})$, and the one-to-one correspondence between self-adjoint operators and unitary one parameter groups is then a corollary to Proposition 2.81. Two other important Lie groups whose elements are operators are invertible pseudo-differential operators and invertible Fourier integral operators.[133]

Example 2.105. Gauge groups form another example of infinite dimensional Lie groups that are of great importance in modern physics. The group elements of a gauge group are functions in $\mathcal{F}(\mathcal{M})$ on a finite dimensional manifold $\mathcal{M}$, and the group operations are addition of functions $f \circ g = f + g$ and $\mathrm{Inv}(f) = -f$, and the identity element is the zero section. Since addition

---

[133]Interestingly, Fourier integral operators are a principal bundle over the diffeomorphism group of canonical transformations.

commutes, the group is Abelian, and since it is a vector group its Lie algebra is $\mathfrak{g} \cong T_e G \cong \mathcal{F}(\mathcal{M})$. The above gauge group can be generalized by replacing addition by multiplication, and more generally by any finite dimensional Lie group; this leads for example to the notion of a Loop group when addition is replaced by the circle (group) S^1.

Diffeomorphism Groups   The group $\mathrm{Diff}(\mathcal{M})$ of a finite dimensional manifold $\mathcal{M}$ is formed by the set of all diffeomorphisms of $\mathcal{M}$ with composition as group multiplication. Invertibility is guaranteed by the definition of a diffeomorphism, cf. Def. 2.62, and one usually restricts oneself to orientation preserving diffeomorphisms which corresponds to the connected component of $\mathrm{Diff}(\mathcal{M})$ containing the identity. The set of smooth diffeomorphisms can be equipped with a manifold structure, but it is a priori neither Banach nor Hilbert but only Fréchet. The usual approach to obtain a better behaved structure is to complete the smooth diffeomorphisms $\mathrm{Diff}^\infty(\mathcal{M})$ is the C^k Banach norm or in a H^s Sobolev norm for $s > 1/2 \dim(\mathcal{M})$.[134] Using inverse limits, as introduced in Def. 2.181, one can define inverse limit Banach Lie groups (ILB) and inverse limit Hilbert Lie groups (ILH) using the above completions that extend to the smooth case.[135] In the literature, usually $\mathrm{Diff}^s(\mathcal{M})$ based on the completing in H^s is considered since it affords locally the conveniences of a Hilbert space setting. However, even then the group operations are not smooth but one has for multiplication and the inversion map that

$$\mathrm{Diff}^{s+k}(\mathcal{M}) \times \mathrm{Diff}^s(\mathcal{M}) \to \mathrm{Diff}^s(\mathcal{M})$$

$$\mathrm{Inv}(\mathrm{Diff}^{s+k}(\mathcal{M})) \to \mathrm{Diff}^s(\mathcal{M})$$

and in particular when $k = 0$ and two diffeomorphisms of the same Sobolev order s are employed then the group operations are only continuous. The standard construction assumes that $\mathcal{M}$ is compact but it was recently shown[136] how to construct such groups when $\mathcal{M}$ is not compact but of bounded geometric complexity, which is the setting of physical relevance.

The Lie algebra $\mathfrak{g}$ of the diffeomorphism group $\mathrm{Diff}(\mathcal{M})$ is naturally identified with $(\mathfrak{X}(\mathcal{M}), [,])$, that is with the space of vector fields $\mathfrak{X}(\mathcal{M})$ on $\mathcal{M}$ under the negative Jacobi-Lie bracket $[,]$: Consider a curve $g(t) : [a, b] \to \mathrm{Diff}(\mathcal{M})$ on $\mathrm{Diff}(\mathcal{M})$. Then for $\bar{x} \in \mathcal{M}$ the diffeomorphism generates a curve $g(\bar{x}, t) : [a, b] \to \mathcal{M}$ at $\bar{x}$ whose tangent vectors are elements in $T_{g(\bar{x},t)}\mathcal{M}$. Since a diffeomorphism $\mathrm{Diff}(\mathcal{M})$ acts simultaneously on all $x \in \mathcal{M}$ one vector element in the Lie algebra is a vector field in $\mathfrak{X}(\mathcal{M})$ on $\mathcal{M}$. That the Lie bracket has to be the negative Jacobi-Lie bracket for vector fields follows from Proposition 2.83

---

[134]The requirement $s > 1/2 \dim(\mathcal{M})$ is necessary for the Sobolev embedding theorem to become available, and only then is the manifold structure chart independents.

[135]Omori ("On the group of diffeomorphisms on a compact manifold") introduced the concept of inverse limit groups, and Ebin and Marsden ("Groups of Diffeomorphisms and the Motion of an Incompressible Fluid") developed it into the form still widely used today.

[136]Eichhorn and Schmid, "Form preserving diffeomorphisms on open manifolds"; Eichhorn, "Partial differential equations on closed and open manifolds".

since we can think of $\mathrm{Diff}(\mathcal{M})$ as the action $G \times \mathcal{M} \to \mathcal{M}$ of some group G on $\mathcal{M}$, in which case the Lie group homomorphism implies a Lie algebra homomorphism, cf. also Remark 2.142. The exponential map is generated by the flow of vector fields on $\mathcal{M}$, but it is not a local diffeomorphism and not even locally surjective so that there are diffeomorphisms arbitrary close to the identity that are not generated by vector fields.

Example 2.106. The diffeomorphism group on space-time is central to the general theory of relativity where it describes invariance under coordinate transformations, and is just a more modern and technical re-statement of Einstein's original postulate of covariance, cf. the change of variables theorem for manifolds in Theorem 2.21. With this symmetry, the Einstein field equations are Hamiltonian on the space of all metrics modulo C^∞ diffeomorphisms.[137]

Example 2.107. The volume preserving diffeomorphism group $\mathrm{Diff}_\mu(\mathcal{M})$ is the Lie subgroup of all diffeomorphisms $\mathrm{Diff}(\mathcal{M})$ that preserve the volume form, or, more generally, a one-density on $\mathcal{M}$, and it inherits its structure from the general diffeomorphism group. The Lie algebra of Diff_μ is the space $\mathfrak{X}_{\mathrm{div}}(\mathcal{M})$ of all divergence free vector fields, again together with the negative Jacobi-Lie bracket, which is as an immediate consequence of the definition of divergence freeness. The volume preserving diffeomorphism group is the configuration space of the ideal Euler fluid, and time evolution of a fluid can be described as a geodesic on the group. We will develop this example in more detail in Example 2.147. The volume preserving diffeomorphism group also plays an important role in optimal transport theory.[138]

Example 2.108. The symplectomorphism group $\mathrm{Diff}_{\mathrm{can}}(\mathcal{M})$, the group of canonical transformation preserving the symplectic 2-form, is another Lie subgroup of $\mathrm{Diff}(\mathcal{M})$, and it is of great importance for Hamiltonian mechanics. We will discuss this group in more detail in the following section after we introduced the necessary concepts from Hamiltonian mechanics.

Remark 2.146. As discussed before, diffeomorphisms are isomorphism in the category of smooth manifolds. Diffeomorphism groups thus arise naturally in many applications, and in particular in geometric mechanics where the configuration space of dynamical systems is a manifold. For many systems one thinks of a test particle in a vector field on configuration or phase space that realizes only a single curve. However, the vector field is usually defined at least in some neighborhood and one hence has at least a local diffeomorphism of some neighborhood generated by the vector field; we previously used flow boxes to formalize this idea. Diffeomorphism groups are hence central to continuum mechanics where, at least intuitively, the limit of an infinite number of particles is studied and one thus no longer considers the flow along individual trajectories

---

[137]Schmid, "Infinite-Dimensional Hamiltonian Systems".

[138]See for example (Khesin et al., "Geometry of diffeomorphism groups, complete integrability and optimal transport").

but of entire regions. On the level of the group, however, the time evolution of the system is again described by a single curve.

Remark 2.147. The use of representations of infinite dimensional Lie groups on function space, as pioneered in the work by Koopman, Stone, von Neumann, and somewhat later van Hove, cf. Chapter 2.3.2.7 and in particular Remark 2.129, is one approach that enables to avoid many of the technicalities of infinite dimensional groups. Then, one only has to work with operators on infinite dimensional spaces. But these are well understood and, moreover, often chosen to be unitary which further facilities their treatment.

2.3.4 Geometric Hamiltonian Mechanics[139]

In the following, we will present the modern formulation of Hamiltonian mechanics. Calculus on manifolds will be employed from the outset, which intrinsically ensures covariance and exposes the structure of the theory, and for intuition our primer on geometric mechanics in Chapter 2.3.1 should be kept in mind throughout.[140] Canonical systems will be considered first, both in the finite and infinite dimensional case, and in the next section, Chapter 2.3.5, systems with symmetries are discussed. However, before the mathematics of Hamiltonian mechanics and its geometry are considered some remarks on the nature of mechanics are useful.

2.3.4.1 Mechanics

In the following, we will briefly remark on what constitutes mechanics—for the purpose of the present thesis—and what are the questions being investigated, and the conceptual perspective of this section will provide context for the remainder of the chapter.

Classically, mechanics was concerned with the dynamics of masses and continua.[141] From a contemporary point of view, the following, more general definition of the subjects of mechanics is warranted.

---

[139]Principal references for the material in this section are (Abraham and Marsden, *Foundations of Mechanics*; Arnold, *Mathematical Methods of Classical Mechanics*; Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*; Holm, Schmah, and Stoica, *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*; Ratiu, "A Crash Course in Geometric Mechanics"). Our presentation differs from the literature only in that we include many worked examples, which hopefully will help the reader to gain an understanding on how the theory is employed in practice and should also provide further assistance with the mathematics developed in the foregoing. Proofs in this section are also include for this purpose, and we refer to the aforementioned references for missing results.

[140]Expositions of the "bare hands" theory on Euclidean space, but geared towards a geometric treatment, can for example be found in (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, Chapter 2) and (Holm, Schmah, and Stoica, *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*, Chapter 1).

[141]See the introductory remarks for Chapter 2.3.
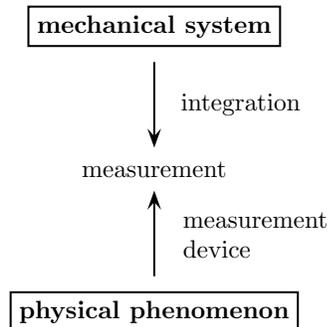
Figure 2.42: Measurements establish the connection between a mechanical system and the modelled physical phenomenon.

Definition 2.197. *A **mechanical system** is a mathematical representation of an observable phenomenon in the physical world.*

The representations that are employed in mechanics are motivated by the desire to qualitatively and quantitatively describe the time evolution of physical phenomena, and the requisite language for this is provided by mathematics. In fact, much of classical mathematics was developed to obtain more effective representations for mechanical systems, and in the foregoing we discussed for example how Lagrangian and Hamiltonian mechanics are mathematical representations for the dynamics of a mechanical system on configuration and phase space, respectively.[142]

Often neglected in mechanics are measurements—despite the fact that they provide the only verifiable facet of a mechanical system, and are hence fundamental for establishing the connection between the mathematical models and the physical world, see Fig. 2.42.[143]

Definition 2.198. *An **observable** of a mechanical system is a variable whose integral provides a **measurement** that can be reproduced in the physical world using a measurement device.*

Our conceptualization of a measurement represents both the integral of an observable variable in the mathematical model and the corresponding aspect of the physical phenomenon modelled by the mechanical system, and it is to be understood as a possibility of effectuation which is independent of any technical

---

[142]Mechanics is concerned with *time* evolution and time plays an intricate rule in the theory where it appears at once as an independent dimension but at the same time is woven into the mathematical and physical structure of a system (which becomes very apparent when the continuity equation for a function is considered in space-time). The topic has so far not received the necessary attention in the literature. Some preliminary remarks can be found in (Belot, "The Representation of Time and Change in Mechanics").

[143]The neglect of observables is particularly frappant in the theoretical mechanics literature, and in contrast to quantum mechanics where these play a central role. A notable exception is (Sudarshan and Mukunda, *Classical Dynamics: A Modern Perspective*).

realization. Measurements are hence the realizations of observables, and what is described by the equations embodying the mechanical system is the time evolution of the observables. However, observables alone are not necessarily sufficient to describe time evolution, and usually also dynamical variables such as velocity are needed. For example, the fluid velocity in hydrodynamics, whose time evolution is described for example by Euler's ideal fluid equations, is a dynamical variable, while the mass density is an observable, and the associated measurement is the mass flowing through a surface in a finite time interval. Analogously, the angular momentum is the dynamical variable for the Euler top, while the moment of inertia tensor is the observable, which is the collective mass density obtained by integration in the body reference frame, and a measurement can be understood as the mass in a local volume of the body as described by the inertia tensor.[144]

Remark 2.148. The distinction between measurements and observables can be obscure at times, in particular when implicit assumptions about a physical phenomenon are made. For instance, flux through a surface over a finite time interval is a measurement while flux through a surface is an observable, cf. Remark 2.114. However, for a stationary system the numerical difference is only a factor proportional to the measurement time, and since in practice almost any system can be assumed to be quasi-stationary over a sufficiently small time interval no distinction is often made at all between surface flux and integrated surface flux. Nonetheless, when physical measurements are performed the difference is still of importance.

Remark 2.149. The distinction between measurements and observables is vital for obtaining effective mechanical systems by providing flexibility in the mathematical representation that is employed: a phenomenon is bound to a set of measurables but observables and dynamical variables are modeling choices that are in principle suitable as long as measurements agree with physical observation. The resulting non-objective nature of observables is apparent in the different scales existing in physics. For example, modeling a fluid as a continuum is absurd—a fluid consists of a large number of discrete and distant molecules. Nonetheless, the fluid velocity and mass density, which are the dynamical variable and observable resulting from the continuum assumption, together with the Euler fluid and Navier-Stokes equations governing their time evolution, are highly effective in describing the macroscopic behaviour of incompressible liquids.

In Def. 2.197 we stipulated a mechanical system as a mathematical model describing a phenomenon in the physical world. In the light of the preceding discussion on measurements and observables we will additionally require the

---

[144]Cf. (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, Chapter 15).

following three principles for a mechanical system to be well defined and provide an effective representation:[145]

Covariance, we expect a phenomenon to be independent of any coordinate system that might be employed for its physical description.

Covariance under differentiation, we expect a mechanical system to be coordinate independent also when it is described by differential equations.

Measurements are integrals, measurements, the quantities of intrinsic physical significance, are obtained by integrating observables and these are again independent of a coordinate system that might be employed.

A constitutional corollary of the above principles of mechanics is that **observables in classical mechanics are differential forms**. Covariance requires observables and dynamical variables to be well behaved under coordinate transformation, covariance under differentiation requires an anti-symmetric derivative, and measurements require observable to be integrands. Differential forms are tensors, and hence intrinsically well defined under coordinate transformations, the exterior derivative defined on forms is anti-symmetric and the covariant part of the usual derivative,[146] and differential forms are " ... ready (or designed, if you prefer) to be integrated ...".[147]

Remark 2.150. In the physics and engineering literature, an emphasis is sometimes placed on units, and the physical plausibility of an equation is deduced from the consistency of the units employed.[148] However, units are not intrinsic, which is probably best illustrated by artificial ones such as steradians, and the validity of a mechanical system should be judged by its internal consistency rather than artificial concepts. With differential forms, the units of an observable are always those of the measurement that is obtained when these are integrated, such as those of mass or energy,[149] and verifying the consistency of a mathematical model is facilitated by the rich structure provided by the exterior calculus. For example, vectors and differential forms of different degrees are distinguished, and operations such as the wedge product of a vector and a differential form are not meaningful. Additionally, the basis functions of differential forms can serve as a substitute for units which is intrinsically defined, for example writing "$F\,dt$" for the flux through a surface expresses important characteristics of the quantity under consideration.

---

[145]These principles are surely well known in the literature although we are not aware of works where these are spelled out explicitly.

[146]That every differential equation can be formulated using the exterior derivative is shown for example in (Sharpe, *Differential geometry: Cartan's generalization of Klein's Erlangen program*).

[147]Desbrun, Kanso, and Tong, "Discrete Differential Forms for Computational Modeling", p. 3.

[148]See also (Arvo, "Transfer Equations in Global Illumination").

[149]Cf. (Deschamps, "Electromagnetics and Differential Forms").

After establishing necessary foundational notions on mechanics, we are prepared to develop its mathematical formulation and geometric structure. We will begin by developing a quasi-classical formulation of Hamiltonian mechanics which is suited for a geometrization but still rooted in computational formulas. Section 2.3.4.3 will then develop the geometry that underlies the theory.

2.3.4.2 Hamiltonian Mechanics

In Chapter 2.3.1 we discussed the space of all possible configurations for various mechanical systems and observed that these are described by manifolds. This motivates the following postulate.[150]

Postulate 2.1. *The* **configuration space** Q *of a mechanical system is a manifolds, and the system's* **configurations** *are described by points on* Q.

Examples for configuration spaces were provided in Table 2.1 and it is important to keep in mind that a point $q \in Q$ on configuration space represents all variables that are relevant for the dynamics of a system, which are not necessarily only classical configuration parameters such as position and orientation but can also encompass variables such as magnetization or chemical density. Postulate 2.1 also includes systems where the configuration space Q is provided by a Lie group with its manifold structure, as for example in the case of the Euler top or the ideal Euler fluid. For finite dimensional systems we have in addition to the foregoing postulate the following.

Definition 2.199. *When the configuration space* Q *of a mechanical system is finite dimensional, the configurations are described by* **generalized coordinates** $q = (q^1, \ldots, q^n)$.

For the examples considered in Table 2.1 the time evolution was described by curves on configuration space. This holds for a much large class of mechanical systems and motivates the following postulate.[151]

Postulate 2.2. *The time evolution of a mechanical system is described by a curve* $q(t) : [a, b] \subset \mathbb{R} \to Q$ *on configuration space* Q.

The smoothness requirement in the above postulate excludes many interesting systems such as those having impacts. However, it suffices for our purposes and greatly simplifies the subsequent mathematical treatment. The curve describing the time evolution of a mechanical system is sometimes referred to as the system's path or trajectory.

---

[150]Using a refined definition of a mechanical system, it might be possible to transform the postulate into a theorem, but for our purposes the present form suffices.

[151]Again, with suitable assumptions it might be possible to transform the postulate to a theorem.

Lagrangian mechanics   Classically, the Hamiltonian formulation of a mechanical system is obtained from its Lagrangian description using the Legendre transform. We will hence in the following briefly discuss Lagrangian mechanics.

Definition 2.200. *The **velocity of a mechanical system** $\dot{q}(t)$ is the tangent vector of the system's trajectory $q(t)$,*

$$\dot{q}(t) = \frac{dq(t)}{dt} \in T_{q(t)}Q.$$

The velocity $\dot{q}(t)$ is an element in the tangent space $T_{q(t)}Q$ at the point $q(t) \in Q$ where the system is at time t, see again Fig. 2.11. The tangent bundle TQ with coordinates $(q, \dot{q})$ hence plays an vital role in the description of the dynamics of a system and it is sometimes denoted as **extended configuration space** or **velocity phase space**.

Definition 2.201. *The **Lagrangian** $L(q, \dot{q}) : TQ \to \mathbb{R}$ of a mechanical system*

$$L = T - V$$

is the kinetic energy T minus the potential energy V.

The Lagrangian can be regarded as a characteristic function for a mechanical system, and its importance arises from Hamilton's principle which, as discussed before, intuitively defines paths of "least resistance", see again Fig. 2.14. The "resistance" of a path on configuration space if formalized as follows.

Definition 2.202. *Let Q be the configuration space for a mechanical system with Lagrangian L, and let $\mathcal{P}(Q)$ be the space of all paths $q(t) : [a, b] \to Q$. The **action functional** $S : \mathcal{P}(Q) \to \mathbb{R}$ is then*

$$S\left(q(t)\right) = \int_{t_0}^{t_1} L\left(q(t), \dot{q}(t)\right) dt. \tag{2.191}$$

With the action functional, we can formally introduce Hamilton's principle of least action.

Postulate 2.3 (Hamilton's action principle)**.** *Let Q be the configuration space for a mechanical system with Lagrangian L. Then physical paths $q(t)$ of the system on configuration space are stationary points*

$$0 = \frac{\delta S}{\delta q(t)} = \frac{d}{d\epsilon}S(q(t) + \epsilon\, r(t))$$

of the action functional $S(q(t))$ for fixed endpoints $q(t_0)$ and $q(t_1)$.

The fixed endpoints correspond to the initial and final configurations of the system, and Hamilton's action principle determines physical motions $q(t)$ that take the system from $q(t_0)$ to $q(t_1)$. In the form employed in Postulate 2.3, the functional derivative is often referred to as a variation δS of the action and the extrema are studied in the calculus of variations, although it is insightful

to keep in mind that δS is just the usual functional derivative on path space. The action principle provides a global description of the physical motions of a system that furnishes much insight into the structure of the dynamics. A local, differential description of the dynamics is provided by the second order **Euler-Lagrange equation** which is given by

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0. \tag{2.192}$$

As shown in Remark 2.32, the equation can be derived from Hamilton's principle using the calculus of functional derivatives.

Example 2.109. The Lagrangian for a classical particle of mass m in a potential $V : Q \to R$ is given by

$$L(q, \dot{q}) = \frac{m}{2}\|\dot{q}\|^2 - V(q) \tag{2.193}$$

where the kinetic energy is

$$T = \frac{m}{2}\|\dot{q}\|^2. \tag{2.194}$$

Evading Hamilton's action principle and directly employing the Euler-Lagrange equations in Eq. 2.192 for the above Lagrangian yields

$$0 = \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} \tag{2.195a}$$

$$= \frac{d}{dt}\frac{\partial}{\partial \dot{q}^i}\left(\frac{m}{2}\|\dot{q}\|^2\right) - \frac{\partial}{\partial q^i}V(q) \tag{2.195b}$$

$$= \frac{d}{dt}\left(\frac{2m}{m}\|\dot{q}\|\frac{\dot{q}}{\|\dot{q}\|}\right) - \nabla V(q) \tag{2.195c}$$

$$= m\,\ddot{q} - \nabla V(q). \tag{2.195d}$$

When the gradient $\nabla V(q)$ of the potential is identified with force F and the usual expression for the momentum is used, $p = m\,\dot{q}$, this is just Newton's second law

$$F = \dot{p} = m\ddot{q}, \tag{2.196}$$

stating that the change in the momentum p of the particle is given by the force F acting on it. Assuming the particle possesses charge e and $V(q)$ is given by the classical Coulomb potential

$$V(q) = e\,\frac{1}{4\pi}\frac{1}{\|q\|} \tag{2.197}$$

relative to a point charge at the origin, the equations of motion become

$$m\,\ddot{q} = e\,\frac{1}{4\pi}\frac{\bar{q}}{\|q\|^2} \tag{2.198}$$

with $\bar{q}$ being a unit vector and we assumed units to be geometrized. The above equation is the classical expression for the dynamics of a particle in an electric field.

Legendre Transform With the Lagrangian description of a mechanical system on configuration space Q, its Hamiltonian formulation on the cotangent bundle T^*Q is obtained using the Legendre transform. The momentum, the analogue of the velocity on T^*Q, is obtained using fiber derivative that was introduced in Def. 2.99.

Definition 2.203. *The **canonical momentum** $p \in T^*Q$ of a mechanical system with configuration space Q and Lagrangian L is the fiber derivative*

$$p = \mathbb{F}L,$$

or in coordinates,

$$(q^i, p_i) = \left(q^i, \frac{\partial L}{\partial \dot{q}^i} \right).$$

It is important to realize that the canonical momentum is an intrinsic property of a mechanical system, and not an independent concepts such as velocity. This is also illustrated by the following examples.

Example 2.110. Consider again the classical particle with the Lagrangian given in Eq. 2.193. By Def. 2.203, we then have for the canonical momentum

$$p = \frac{\partial L}{\partial \dot{q}} = \frac{\partial}{\partial \dot{q}} \left(\frac{m}{2} \|\dot{q}\|^2 - V(q) \right) = \frac{2m}{2} \|\dot{q}\| \frac{\dot{q}}{\|\dot{q}\|} = m \, \dot{q} \tag{2.199}$$

which is Newton's classical expression for the momentum of a particle.

Example 2.111. The Lagrangian for a charged particle in an electromagnetic field is

$$L(q, \dot{q}) = \frac{m}{2} \|\dot{q}\|^2 - e \, \varphi + \frac{1}{c} \dot{q} \cdot A \tag{2.200}$$

where e is the particle's electrical charge, φ is the electric charge density, and A is the magnetic vector potential. By Def. 2.203, the canonical momentum of the particle is hence

$$p = \frac{\partial L}{\partial \dot{q}} = \frac{\partial}{\partial \dot{q}} \left(\frac{m}{2} \|\dot{q}\|^2 - \varphi + \frac{1}{c} \dot{q} \cdot A \right) = m \, \dot{q} + A. \tag{2.201}$$

The expression for the momentum in Eq. 2.201 should be compared to those in Eq. 2.199, which shows that the momentum is *not* a property of a particle but of an entire mechanical system.

Remark 2.151. A word of caution regarding the concept of momentum is in order at this point. In the literature, momentum refers to a variety of related but nonetheless distinct concepts. We saw for example how the canonical momentum is defined through the fiber derivative. In contrast, the angular momentum of the Euler top arises from its symmetries and is best understood using the concept of a momentum map, cf. Chapter 2.3.5.

Next to the momentum, the fiber derivatives also enables to define the energy of a mechanical system, which will allow us in the following to introduce the Hamiltonian function of a system.

Definition 2.204. *The **energy function** $E : TQ \to \mathbb{R}$ of a mechanical system with configuration space Q and Lagrangian L is*

$$E(q, \dot{q}) = \mathbb{F}L(\dot{q}) \cdot \dot{q} - L(q, \dot{q}) = p \cdot \dot{q} - L(q, \dot{q}). \tag{2.202}$$

Intuitively, the energy function represents the energy of a system moving with velocity $(q, \dot{q}) \in T_q Q$ This is illustrated in the following example.

Example 2.112. The Lagrangian for a classical particle was introduced in Eq. 2.193. With the expression for the momentum from Example 2.110, we obtain for the energy of the system

$$E(q, \dot{q}) = p \cdot \dot{q} - L(q, \dot{q}) \tag{2.203a}$$

$$= m \, \dot{q} \cdot \dot{q} - \left(\frac{m}{2} \|\dot{q}\|^2 - V(q) \right) \tag{2.203b}$$

$$= m \, \|\dot{q}\|^2 - \frac{m}{2} \|\dot{q}\|^2 + V(q) \tag{2.203c}$$

$$= \frac{m}{2} \|\dot{q}\|^2 + V(q) \tag{2.203d}$$

which is the expression to be expected from the definition of the Lagrangian in Def. 2.201.

With the momentum, the second step of the Legendre transform is to "pass" the characteristic function of the system from TQ to T^*Q. However, this is only possible when the mechanical system is sufficiently "well behaved".

Definition 2.205. *The Lagrangian L of a mechanical system is **hyperregular** when the fiber derivative $\mathbb{F}L$ is a diffeomorphism from the tangent bundle TQ to the cotangent bundle T^*Q.*

Since the Lagrangian is characteristic for a system, one can speak of hyperregular systems, and we will restrict ourselves to such systems in the following. From the definition of the momentum in Def. 2.203, it also follows that for a hyperregular system momentum and velocity of a system are in one-to-one correspondence, and it makes intuitively sense that this is a necessary condition to pass the dynamics of a system from TQ to T^*Q. On the cotangent bundle T^*Q the generator of the dynamics is no longer the Lagrangian but the Hamiltonian.

Definition 2.206. *The **Hamiltonian function** or **Hamiltonian** $H : T^*Q \to \mathbb{R}$ of a mechanical system system with configuration space Q and Lagrangian L is*

$$H = E \circ (\mathbb{F}L)^{-1}$$

where E is the system's energy function, and more explicitly

$$H(q,p) = p \cdot \dot{q} - L(q, \dot{q}).$$

*A **Hamiltonian system** is the triple (T^*Q, H, p) and T^*Q is then referred to as **phase space**.*

By definition, the Hamiltonian is the pre-image of a system's energy on the cotangent bundle, and it is through this indirection that some scepticism is in order when the Hamiltonian is interpreted too naïvely as the total energy of a system.

Example 2.113. Returning to the example of a classical particle that was considered before, we have by Example 2.110 for the fiber derivative

$$\mathbb{F}L: \ TQ \to T^*Q \ : \ \dot{q} \to p = m\,\dot{q} \tag{2.204a}$$

and its inverse is hence

$$(\mathbb{F}L)^{-1}: \ T^*Q \to TQ \ : \ p \to \dot{q} = \frac{p}{m}. \tag{2.204b}$$

It is easy to see that these maps provide a diffeomorphism between TQ and T^*Q as long as $m > 0$, ensuring that the system is hyperregular. With the expression for the energy in Example 2.112, we thus obtain for the system's Hamiltonian

$$H(q,p) = E(q, \dot{q}) \circ (\mathbb{F}L)^{-1} \tag{2.205a}$$

$$= \left(\frac{m}{2} \|\dot{q}\|^2 + V(q) \right) \circ \left(\dot{q} \mapsto \frac{p}{m} \right) \tag{2.205b}$$

$$= \frac{m}{2} \left\| \frac{p}{m} \right\|^2 + V(q) \tag{2.205c}$$

$$= \frac{1}{2m} \|p\|^2 + V(q) \tag{2.205d}$$

which is indeed the expression found in the literature. Using the more explicit expression for the Hamiltonian in terms of the Lagrangian we obtain

$$H(q,p) = p \cdot \dot{q} + L(q, \dot{q}) \tag{2.206a}$$

$$= p \cdot \frac{\dot{p}}{m} - \left(\frac{1}{2m} \|p\|^2 - V(q) \right) \tag{2.206b}$$

$$= \frac{1}{2m} \|p\|^2 + V(q). \tag{2.206c}$$

It should be observed that using the inverse fiber derivative $(\mathbb{F}L)^{-1}$ is both computationally simpler and conceptually more insightful.

For a hyperregular system the tangent and cotangent bundle are in bijective correspondence and hence the definition of the momentum can be "inverted",

$$\dot{q} = \frac{\partial H}{\partial p}. \tag{2.207}$$

Dynamics of Hamiltonian Systems A Hamiltonian system provides a complete description of a mechanical system on the cotangent bundle. What is lacking so far, however, is the equivalent of the Euler-Lagrange equations for the description of the dynamics on phase space.

Theorem 2.29. *The velocity of the phase space trajectory of a hyperregular Hamiltonian system is the* **Hamiltonian vector field** $X_H \in T(T^*Q)$ *defined by* **Hamilton's equations** *as*

$$X_H = \begin{pmatrix} \dfrac{\partial q}{\partial t} \\ \dfrac{\partial p}{\partial t} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial H}{\partial p} \\ -\dfrac{\partial H}{\partial q} \end{pmatrix} \in T(T^*Q).$$

Proof. The theorem is established by showing the equivalence to the Euler-Lagrange equations. Indeed, with Eq. 2.207, $\dot{q} = \partial q/\partial t = \partial H/\partial p$ we immediately obtain

$$\begin{pmatrix} \dfrac{\partial q}{\partial t} \\ \dfrac{\partial p}{\partial t} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial q}{\partial t} \\ -\dfrac{\partial H}{\partial q} \end{pmatrix} \tag{2.208a}$$

and the definition of the momentum in Def. 2.203 yields

$$\begin{pmatrix} \dfrac{\partial q}{\partial t} \\ \dfrac{d}{dt}\dfrac{\partial L}{\partial \dot{q}} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial q}{\partial t} \\ -\dfrac{\partial H}{\partial q} \end{pmatrix}. \tag{2.208b}$$

Using the definition of the Hamiltonian we have

$$\frac{\partial H}{\partial q} = \frac{\partial}{\partial q}(p{\cdot}\dot{q} - L) = \frac{\partial L}{\partial q}$$

and hence

$$\begin{pmatrix} \dfrac{\partial q}{\partial t} \\ \dfrac{d}{dt}\dfrac{\partial L}{\partial \dot{q}} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial q}{\partial t} \\ -\dfrac{\partial L}{\partial q} \end{pmatrix} \tag{2.208c}$$

where the first line of the equation is just an identity and the second is the Euler-Lagrange equation. ∎

The Hamiltonian vector field defines trajectories in phase space along the level sets of the Hamiltonian, as was illustrated in Fig. 2.15.

Example 2.114. For our running example of a classical particle, the Hamiltonian was obtained in Example 2.113. The Hamiltonian vector field for the system is hence

$$\frac{\partial q}{\partial t} = \frac{\partial}{\partial p}\left(\frac{1}{2m}\|p\|^2 + V(q)\right) = \frac{2}{2m}\|p\|\frac{p}{\|p\|} = \frac{p}{m} \tag{2.209a}$$

$$-\frac{\partial p}{\partial t} = \frac{\partial}{\partial q}\left(\frac{1}{2m}\|p\|^2 + V(q)\right) = \frac{\partial V(q)}{\partial q} = \nabla V \tag{2.209b}$$

which is again Newton's second law, this time obtained from a Hamiltonian formulation.

Example 2.115. The Hamiltonian for the single pendulum is given by

$$H(\theta, p) = \frac{p^2}{2m\,L^2} - m\,g\,L\cos\theta \tag{2.210a}$$

where the angle θ represents the generalized coordinate for the position, cf. Table 2.3.1. Hamilton's equations for the pendulum are hence

$$\frac{\partial \theta}{\partial t} = \frac{\partial H}{\partial p} = \frac{p}{m\,L^2} \tag{2.211a}$$

$$-\frac{\partial p}{\partial t} = \frac{\partial H}{\partial \theta} = m\,g\,L\sin\theta \tag{2.211b}$$

A result of central importance in mechanics is the conservation of the Hamiltonian along the flow of a Hamiltonian vector field. By the definition of H, this is usually equivalent to the well known law of conservation of energy.

Proposition 2.89. *The flow of a Hamiltonian system along the Hamiltonian vector field X_H preserves the system's Hamiltonian H.*

Proof. Conservation of the Hamiltonian along the flow is equivalent to a vanishing Lagrangian derivative dH/dt. Computing the total time derivative for a time invariant Hamiltonian we obtain

$$\frac{dH}{dt} = \frac{\partial H}{\partial q}\frac{\partial q}{\partial t} + \frac{\partial H}{\partial p}\frac{\partial p}{\partial t} \tag{2.212a}$$

With the definition of the Hamiltonian vector field this becomes

$$\frac{dH}{dt} = \frac{\partial H}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial H}{\partial p}\frac{\partial H}{\partial q} \tag{2.212b}$$

which immediately yields the result $dH/dt = 0$. $\qquad\blacksquare$

Another important property of Hamiltonian systems is an immediate consequence of the definition of the Hamiltonian vector field X_H.

Theorem 2.30 (Liouville's theorem)**.** *The flow of a Hamiltonian system along the Hamiltonian vector field is volume preserving.*

Proof. We saw in Chapter 2.3.1 that Hamilton's equation can be written as

$$X_H = \mathbb{J}\,\nabla H = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}\begin{pmatrix} \nabla_q\,H \\ \nabla_p\,H \end{pmatrix}. \tag{2.213}$$

The symplectic matrix $\mathbb{J}$ can be considered as the generator of the Hamiltonian flow. But its determinant is unity, that is $\det(\mathbb{J}) = 1$, which immediately implies the result. $\qquad\blacksquare$

The last three theorems were proved using a "bare hands" approach with only limited insight into the fabric underlying the results. We will now turn to the geometric formulation of Hamiltonian mechanics where these and additional results follow intrinsically from the geometry.

2.3.4.3 The Geometry of Hamiltonian Mechanics

In the foregoing we developed Hamiltonian mechanics assuming an existing Lagrangian description of a mechanical system. However, the theory can also be formulated intrinsically and this will be the approach taken in this section.

Symplectic manifolds The fabric of geometric Hamiltonian mechanics is provided by symplectic manifolds, and these afford the category for the theory.[152] The definition of a symplectic manifold requires the notion of weakly and strongly non-degenerate 2-forms, which extend the weakly and strongly non-degenerate pairings between vector spaces which were introduced in Chapter 2.2.1. Recall that a pairing of two vector spaces is strongly non-degenerate when it provides an isomorphism between the spaces, and it is weakly non-degenerate when it is one-to-one but not surjective.

Definition 2.207. *A 2-form $\beta \in \Omega^2(\mathcal{M})$ on a manifold $\mathcal{M}$ is **strongly non-degenerate** when*

$$TQ \to T^*Q \ : \ X \mapsto \alpha \ := \ i_X\beta = \alpha$$

*is one-to-one and onto for all $X \in \mathfrak{X}(\mathcal{M})$ and $\alpha \in \mathfrak{X}^*(\mathcal{M}) \cong \Omega^1(\mathcal{M})$. A 2-form is **weakly non-degenerate** when it defines a pairing as above but which is not necessarily onto.*

By its definition, a non-degenerate 2-form introduces a well defined correspondence between the tangent and cotangent bundle of a manifold, and when the 2-form is strongly non-degenerate then it provides an isomorphism between the fibers T_qQ and T_q^*Q. It should also be kept in mind that in the finite dimensional case every weakly non-degenerate 2-form is also strongly non-degenerate. We are now prepared for the principal notion of a symplectic manifold.

---

[152]An ontological difference between Lagrangian and Hamiltonian mechanics is that for the former no category exists, see however (Cendra, Marsden, and Ratiu, *Lagrangian Reduction by Stages*; Cendra, Marsden, and Ratiu, "Geometric Mechanics, Lagrangian Reduction, and Nonholonomic Systems") for some recent efforts.

Definition 2.208. *A **symplectic manifold** (P, ω) is a manifold P equipped with a closed, weakly non-degenerate **symplectic** 2-**form** $\omega \in \Omega^2(P)$. When the 2-form ω is strongly non-degenerate, then (P, ω) is a **strong symplectic manifold**.*

The symplectic 2-form is the intrinsic analogue of the symplectic matrix $\mathbb{J}$ which was used in the foregoing to describe the dynamics of a Hamiltonian system. We will discuss shortly how it can be employed to obtain an intrinsic definition of the Hamiltonian flow. The following theorem characterizes the local structure of a symplectic manifold.

Theorem 2.31 (Darboux's theorem). *Let (P, ω) be a strong symplectic manifold. Then in a neighborhood of each point $z \in P$ there is a local coordinate chart in which the symplectic 2-form ω is constant.*

In other words, Darboux's theorem assures us that locally all symplectic manifolds of the same dimension "look the same". This should be compared to Riemannian geometry where two manifolds "look locally the same" if and only if their curvature tensors agree, which is a very stringent requirement that is generically never satisfied. Since the local structure of all symplectic manifolds is identical, all interesting geometry for them is global.[153] Note that the manifold in Darboux's theorem can be infinite dimensional, but that strong non-degeneracy of the symplectic 2-form is required. In the finite dimensional case Darboux's theorem also implies the following.

Corollary 2.18. *Let (P, ω) be an m-dimensional symplectic manifold. Then P is even dimensional with $m = 2n$ and there exist **canonical coordinates** $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ where the symplectic 2-form takes the form*

$$\omega = \sum_{i=1}^{n} dq^i \wedge dp_i. \tag{2.214}$$

That any symplectic manifold is even dimensional can be seen easily for the special case of a linear symplectic manifold where the representation of the symplectic 2-form in local coordinates is given by the symplectic matrix $\mathbb{J}$. With I being the $n \times n$ identity matrix, the symplectic matrix is

$$\mathbb{J} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \tag{2.215}$$

but when we require even weak non-degeneracy, then adding a single dimension is not possible.

Example 2.116. An even-dimensional, real vector space equipped with the symplectic matrix $\mathbb{J}$ is a symplectic manifold, and it is a special case of of a symplectic vector space. By Darboux's theorem, every strong symplectic manifold looks locally like a symplectic vector space and for every finite symplectic manifold the symplectic structure can be represented by the symplectic matrix.

---

[153]The global structure of symplectic manifolds is studied in symplectic topology, see for example (McDuff and Salamon, *Introduction to Symplectic Topology*).

Example 2.117. The 2-sphere S^2 is a symplectic manifold whose symplectic form is given by the usual volume form $d\omega = \sin\theta d\theta \wedge d\phi$ on S^2. Indeed, S^2 is even dimensional, ω is closed since it is a volume form, and it is easily seen to be non-degenerate when the sphere is appropriately covered by multiple charts.

Example 2.118. The classical example of a symplectic manifold, and the original model for the concept, is the cotangent bundle T^*Q of an n-dimensional manifold Q, and the space is by construction even dimensional. The symplectic 2-form on T^*Q arises from the **canonical 1-form** $\theta \in \Omega^1(T^*Q)$ which in coordinates is given by[154]

$$\theta = p_i \, dq^i \in \Omega^1(T^*Q). \tag{2.216}$$

The canonical 1-form is a degenerate 1-form on T^*Q in that the components with respect to the fiber basis functions dp^i are zero, and it hence *appears* like a 1-form on configuration space Q although p_i has clearly no meaning there. With the canonical 1-form θ, the **canonical symplectic 2-form** ω **on** T^*Q is defined as the negative exterior derivative of θ,

$$\omega = -\mathrm{d}\theta = q^i \wedge p_i = \sum_{i=1}^{n} q^i \wedge p_i. \tag{2.217}$$

That ω is closed follows immediately from the identity $\mathrm{d} \circ \mathrm{d} = 0$ and non-degeneracy can be established for example by representing ω in local coordinates by the symplectic matrix $\mathbb{J}$.

The cotangent bundle T^*Q is a natural symplectic manifold also when Q is infinite dimensional. The canonical 1-form $\theta \in T^*(T^*Q) \cong \Omega^1(T^*Q)$ is then defined intrinsically as

$$\theta_\beta(X) = \langle \beta, (T\pi_Q)\, X \rangle \tag{2.218}$$

where $\beta \in T^*Q \cong \Omega^1(Q)$, $X \in T_\beta(T^*Q)$, $\pi_Q : T^*Q \to Q$ so that $T\pi_Q : T(T^*Q) \to TQ$ is the tangent of the map, and the pairing on the right hand side is the natural pairing between T^*Q and TQ. Intuitively, this definition can be interpreted as

$$\theta_p(X_H) = \langle p, (T\pi_Q)X_H \rangle = \langle p, \dot{q} \rangle \tag{2.219}$$

which can be recognized as the first part of the definition of the Hamiltonian in terms of the Lagrangian, cf. Def. 2.206, and for a Lagrangian of the form $L(q, \dot{q}) = T(\dot{q}) - V(q)$ this is just a multiple of the systems kinetic energy. Even in the infinite dimensional case, the symplectic 2-form is defined by $\omega = -\mathrm{d}\theta$ since it was a priori an intrinsic definition.

Any finite dimensional symplectic manifold has a natural volume form induced by the symplectic 2-form, and it is hence an orientable manifold.

---

[154]In the literature, the canonical 1-form is sometimes also referred to as tautological 1-form or Liouville 1-form, and the latter one should not be confused with the Liouville form which is the natural volume form on a symplectic manifold, and which will be introduced shortly.

Theorem 2.32. *Every $2n$-dimensional symplectic manifold (P, ω) is a volume manifold with the volume form given by the **Liouville form***

$$\varpi = \omega \wedge \ldots \wedge \omega = \omega^n \in \Omega^n(P).$$

In local coordinates, the Liouville form is given by[155]

$$\varpi = \frac{n!}{(-1)^{n(n-1)/2}} \, dq^1 \wedge \ldots \wedge dq^n \wedge dp_1 \wedge \ldots \wedge dp_n.$$

A final notion of interest in the context of symplectic manifolds are Lagrangian submanifolds, which arise naturally in many physical applications.

Definition 2.209. *Let (P, ω) be a symplectic manifold of dimension $2n$. A **Lagrangian submanifold** L_ω of P is an n-dimensional submanifold such that the restriction $\omega \mid_{L_\omega}$ of the symplectic 2-form ω to L_ω vanishes.*

Remark 2.152. Symplectic structures are ubiquitous in various branches of mathematics and the sciences, and many classical results can be formulated using symplectic geometry. This been called the "symplectification of science".[156]

Canonical transformations and Hamiltonian dynamics   Natural diffeomorphism in the category of symplectic manifolds are those preserving the symplectic 2-form, and we will see shortly that such diffeomorphisms essentially characterize dynamics in Hamiltonian mechanics.

Definition 2.210. *A smooth mapping $\varphi : P_1 \to P_2$ between symplectic manifolds (P_1, ω_1) and (P_2, ω_2) is **canonical** or **symplectic** when*

$$\varphi^* \omega_2 = \omega_1.$$

Canonical transformations also preserve the Liouville form since $\varphi^* \omega^n = \varphi^* \omega \wedge \ldots \wedge \varphi^* \omega$, which follows from the properties of the pullback for differential forms, cf. Proposition 2.59, i). This together with the inverse function theorem also implies that a symplectic transformation provides a local diffeomorphism when the manifolds P_1 and P_2 in the above definition have the same dimension.

Example 2.119. A classical example for a canonical transformation is the cotangent lift of a diffeomorphism on configuration space. Recall from Def. 2.95 that the cotangent lift $T^*\varphi : T^*Q_2 \to T^*Q_1$ of a diffeomorphism $\varphi : Q_1 \to Q_2$ between manifolds Q_1 and Q_2 is defined by

$$\langle (T^*\varphi) \, \alpha, v \rangle = \langle \alpha, (T\varphi) \, v \rangle \tag{2.220}$$

where $\alpha \in T^*_{q_2} Q_2$, $v \in T_{q_1} Q_1$, $T\varphi$ the tangent map of φ, and where $\langle \, , \rangle$ denotes the natural pairing between the tangent and cotangent spaces on the respective

---

[155]The choice of the constant is purely conventional. For computations it is often useful to choose it such that the coordinate expression is as simple as possible.

[156]Gotay and Isenberg, "La Symplectification de la Science".

manifolds. From Eq. 2.220, we can think of the cotangent lift $T^*\varphi$ of φ as the formal adjoint of the tangent map $T\varphi$, and it is often referred to as a "point transformation" since it arises from a diffeomorphism mapping points on Q_1 to points on Q_2. Cotangent lifts are important because they provide the only transformations $T^*Q_2 \to T^*Q_1$ that conserve the canonical 1-form such that $(T^*\varphi)\,\theta_2 = \theta_1$, with θ_1 and θ_2 being the respective canonical 1-forms on T^*Q_1 and T^*Q_2. This also immediately implies

$$(T^*\varphi)\,\theta_2 = \theta_1 \tag{2.221a}$$

$$\mathrm{d}(T^*\varphi)\,\theta_2 = \mathrm{d}\theta_1 \tag{2.221b}$$

and since the pullback and the exterior derivative commute, cf. Proposition 2.59, ii), we obtain

$$(T^*\varphi)\,\mathrm{d}\theta_2 = \mathrm{d}\theta_1 \tag{2.221c}$$

$$(T^*\varphi)\,\omega_2 = \omega_1. \tag{2.221d}$$

so that cotangent lifts are canonical transformations preserving the symplectic 2-form. It should be noted that cotangent lifted dynamics are significantly more restrictive than the "lifting" of dynamics from T^Q to T^*Q we considered before, which is possible for any Lagrangian system.

With the notion of a symplectic transformation we are prepared to generalize Hamilton's equations to symplectic manifolds, and to discuss the rich geometric structure that arises.

Definition 2.211. *Let (P,ω) be a symplectic manifold. A vector field $X \in TM$ is a **Hamiltonian vector field** X_H when there exists a **Hamiltonian** $H : T^*Q \to \mathbb{R}$ such that*

$$\mathrm{i}_{X_H}\omega = \mathrm{d}H.$$

*The triple (P,ω,H) is a **Hamiltonian system** and P the **phase space** of the system. The flow of a vector field X_H is a **Hamiltonian flow** $\varphi_t : \mathbb{R} \times P \to P$ and the space of all Hamiltonian vector fields will be denoted by $\mathfrak{X}_{\mathrm{Ham}}(P)$. When H exists globally one speaks of a **global Hamiltonian vector field**, otherwise of a **local Hamiltonian vector field**. **Hamilton's equations** for $z \in P$ are*

$$\dot{z} = X_H(z)$$

and the velocity of a point on the symplectic manifold under the Hamiltonian flow φ_t is given by the Hamiltonian vector field.

A consequence of the above definition is the following proposition which provides a complete characterization of canonical transformations.

Proposition 2.90. *Let (P,ω) be a symplectic manifold. The flow $\varphi_t : \mathbb{R} \times P \to P$ of a vector field $X \in TP$ on P is canonical with $\varphi_t^*\omega = \omega$ if and only if X is locally Hamiltonian.*

Proof. The symplectic 2-form ω is conserved when its Lagrangian derivative along the flow vanishes. By the definition of the Lie derivative, cf. Def. 2.158, we have

$$\frac{d}{dt}\varphi^*\omega = \varphi^* \pounds_{X_H}\omega \tag{2.222a}$$

and using Cartan's formula yields

$$\frac{d}{dt}\varphi^*\omega = \varphi^* \left(\mathrm{di}_{X_H}\omega + \mathrm{i}_{X_H}\mathrm{d}\omega \right). \tag{2.222b}$$

The second term vanishes since the symplectic 2-form is closed. With the definition of the Hamiltonian vector field we hence have

$$\frac{d}{dt}\varphi^*\omega = \varphi^*\mathrm{dd}H \tag{2.222c}$$

and by the identity $\mathrm{dd} = 0$ one obtains

$$\frac{d}{dt}\varphi^*\omega = 0. \tag{2.222d}$$

which establishes the proposition. $\qquad\qquad\qquad\blacksquare$

Remark 2.153. The flow of a globally Hamiltonian vector field is known as a **symplectic diffeomorphism** or **symplectomorphism**. These will be considered in more detail in Chapter 2.3.5.3, and they play an outstanding role in the group structure of ideal light transport.

Corollary 2.19 (Liouville's theorem)**.** *The Liouville form* $\varpi = \omega^n$ *is conserved along the flow of a Hamiltonian vector field* X_H.

Before we establish further properties of Hamiltonian flows, let us consider two examples that illustrate the foregoing definitions.

Example 2.120. We discussed in Example 2.118 that the cotangent bundle T^*Q over an n-dimensional configuration space Q is the classical example of a symplectic manifold with the symplectic 2-form given by

$$\omega = \sum_{i=1}^n dq^i \wedge dp_i. \tag{2.223}$$

Since the Hamiltonian is a function on phase space we have by the definition of the exterior derivative that

$$\mathrm{d}H = \sum_{i=1}^n \frac{\partial H}{\partial q^i}dq^i + \frac{\partial H}{\partial p^i}dp^i \tag{2.224}$$

and the Hamiltonian vector field X_H also has to satisfy

$$\mathrm{i}_{X_H}\omega = \mathrm{d}H. \tag{2.225}$$

Comparing Eq. 2.223 and Eq. 2.224, and keeping in mind that the symplectic 2-form is anti-symmetric, suggests then that X_H is given by

$$X_H = \sum_{i=1}^{n} \frac{\partial H}{\partial p^i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p^i}. \tag{2.226}$$

Indeed, repeated application of the Leibniz rule for the interior product, cf. Proposition 2.60, and the biorthogonality of the local basis functions for the cotangent bundle yields that for each term of the interior product $i_{X_H}\omega$, considered as the partial contraction $i_{X_H}\omega = \omega(X_H, \cdot)$, we have

$$dq^i \wedge dp^i \left(\frac{\partial H}{\partial p^i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p^i} \,,\, \cdot \right)$$

$$= dq^i \left(\frac{\partial H}{\partial p^i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p^i} \right) \wedge dp^i - dq^i \wedge dp^i \left(\frac{\partial H}{\partial p^i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p^i} \right) \tag{2.227a}$$

where for the right hand side we used the definition of the wedge product, $(\alpha \wedge \beta)(v, w) = \alpha(v) \wedge \beta(w) - \alpha(w) \wedge \beta(v)$. Since $dq^i(\partial/\partial p^i) = 0$ and $dp^i(\partial/\partial q^i) = 0$ we have

$$dq^i \wedge dp^i \left(\frac{\partial H}{\partial p^i} \frac{\partial}{\partial q^i} - \frac{\partial H}{\partial q^i} \frac{\partial}{\partial p^i} \,,\, \cdot \right) = \frac{\partial H}{\partial p^i} dp^i + \frac{\partial H}{\partial q^i} dq^i \tag{2.227b}$$

which are indeed the components of dH in Eq. 2.224. It is easy to see that the "guess" for the Hamiltonian vector field in Eq. 2.226 is equivalent to the definition provided in the previous section in Theorem 2.29 where we also showed the equivalence to the Euler-Lagrange equations of Lagrangian mechanics. Hence, $z = X_H(z)$ and $i_{X_H}\omega = dH$ provide the intrinsic formulation of the dynamics of a Hamiltonian system, and these do agree with the classical notions when the symplectic manifold is a cotangent bundle T^*Q over a finite dimensional configuration space Q.

Example 2.121.[157] Consider again a classical particle, our running example in the previous section, and assume it has charge e and is immersed in a magnetic field B. Using differential forms, the magnetic field is a 2-form[158]

$$B = B_1 \, dq^2 \wedge dq^3 - B_2 \, dq^1 \wedge dq^3 + B_3 \, dq^1 \wedge dq^3 \tag{2.228}$$

and B is divergence free so that

$$i_B(dq^1 \wedge dq^2 \wedge dq^3) = 0. \tag{2.229}$$

---

[157]The example follows (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, Chapter 6.7).

[158]See for example (Deschamps, "Electromagnetics and Differential Forms"; Stern et al., "Geometric Computational Electrodynamics with Variational Integrators and Discrete Differential Forms") and also our discussion in Chapter 3.2.1.

We equip the space $T^*Q \cong \mathbb{R}^3 \times \mathbb{R}^3$ with the *non-canonical* symplectic form

$$\omega = m\,\omega_{can} - \omega_{mag} = m\left(dq^1 \wedge dp^1 + dq^2 \wedge dp^2 + dq^3 \wedge dq^3\right) - \frac{e}{c}B \quad (2.230)$$

where m is the particle's mass, e its charge, c the speed of light, and it should be kept in mind that the magnetic field B is also a 2-form. The Hamiltonian of the system is given by the kinetic energy

$$H(q,p) = \frac{m}{2}\|p\|^2 = \frac{m}{2}\left(p_1^2 + p_2^2 + p_3^2\right). \quad (2.231)$$

and for $\mathrm{d}H$ we hence obtain

$$\mathrm{d}H = m\,p_i\,dp^i \quad (2.232)$$

so that the Hamiltonian vector field has to satisfy

$$m\,p_i\,dp^i = \mathrm{d}H = \mathrm{i}_{X_H}\omega = m\,\mathrm{i}_{X_H}\omega_{can} - \mathrm{i}_{X_H}\omega_{mag}. \quad (2.233)$$

When X_H has components

$$X_H = v^1\frac{\partial}{\partial q^1} + v^2\frac{\partial}{\partial q^2} + v^3\frac{\partial}{\partial q^3} + a^1\frac{\partial}{\partial p^1} + a^2\frac{\partial}{\partial p^2} + a^3\frac{\partial}{\partial p^3} \quad (2.234)$$

the interior products for the two terms of the symplectic 2-form considered individually are for the canonical term

$$\begin{aligned} m\,\mathrm{i}_{X_H}\omega_{can} = m(\quad & v^1 dp^1 - a^1 dq^1 \\ & + v^2 dp^2 - a^2 dq^2 \\ & + v^3 dp^3 - a^3 dq^3) \end{aligned} \quad (2.235)$$

and for the term arising through the magnetic field

$$\begin{aligned} \mathrm{i}_{X_H}\omega_{mag} = \frac{e}{c}(\quad & B_1\,v^2 dq^3 - B_1\,v^3 dq^2 \\ & - B_2\,v^1 dq^3 + B_2\,v^3 dq^1 \\ & + B_3\,v^1 dq^2 - B_3\,v^2 dq^1). \end{aligned} \quad (2.236)$$

Comparing terms with dp^i in Eq. 2.232 and Eqs. 2.235 and 2.236 we have to have $p_1 = v^1$, $p_2 = v^2$, and $p_3 = v^3$ for $\mathrm{d}H = \mathrm{i}_{X_H}\omega$ to hold. Analogously, the terms with dq^i have to satisfy

$$-m\,a^1 - \frac{e}{c}\left(B_2 v^3 - B_3 v^2\right) = 0 \quad (2.237\text{a})$$

$$-m\,a^2 - \frac{e}{c}\left(B_3 v^1 - B_1 v^3\right) = 0 \quad (2.237\text{b})$$

$$-m\,a^3 - \frac{e}{c}\left(B_1 v^2 - B_2 v^1\right) = 0. \quad (2.237\text{c})$$

Using the identification $T(\mathbb{R}^3 \times \mathbb{R}^3) \cong (\mathbb{R}^3 \times \mathbb{R}^3) \times (\mathbb{R}^3 \times \mathbb{R}^3)$, which is available with the standard Riemannian structure on Euclidean space, we have $v = \dot{q}$ and $a = \ddot{q}$. Recognizing the cross product in Eq. 2.237, the three equations can be written in vector form as

$$-m\,\ddot{q} + \frac{e}{c}\,\dot{q} \times B = 0 \tag{2.238}$$

which is the well known Lorentz force law governing the dynamics of a particle in a magnetic field. The example shows a Hamiltonian system defined on a genuine symplectic manifold where the symplectic 2-form is not the canonical one. The expression for the symplectic 2-form should also be compared to Example 2.111, and we refer to the literature for a more rigorous and complete treatment of the surprisingly rich geometric structure of a classical particle in a magnetic field.[159]

Next to the symplectic 2-form, the flow along the Hamiltonian vector field also preserves the Hamiltonian, as one would expect from the discussion and results in the preceding section.

Proposition 2.91. *The flow φ_t along a Hamiltonian vector field preserves the generating Hamiltonian function H.*

Proof. Infinitesimally, the transport of the Hamiltonian along the Hamiltonian vector field is described by the Lie derivative $\pounds_{X_H} H$ and for H to be preserved we have to have $\pounds_{X_H} H = 0$. Using Cartan's formula one obtains

$$\pounds_{X_H} H = \mathrm{i}_{X_H}\mathrm{d}H + \mathrm{d}\mathrm{i}_{X_H} H \tag{2.239a}$$

The interior product $\mathrm{i}_{X_H} H$ vanishes since the Hamiltonian H is a function, and with the definition of the Hamiltonian vector field we obtain

$$\pounds_{X_H} H = \mathrm{i}_{X_H}\mathrm{i}_{X_H}\omega. \tag{2.239b}$$

But applying the interior product twice is equivalent to the complete contraction of the 2-form ω and by the anti-symmetry of the differential form we hence have

$$\pounds_{X_H} H = \omega(X_H, X_H) = 0 \tag{2.239c}$$

which proofs our claim. $\square$

Remark 2.154. Next to the Hamiltonian other functions might be preserved by the flow of the Hamiltonian vector field and these are so called Casimir functions or Casimirs.

Remark 2.155. When the symplectic manifold (P, ω) is infinite dimensional, some care is required on how the above results are to be interpreted. Computationally, in this case the usual derivative has to be replaced by the functional

---

[159]See for example (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, Chapter 6.6 and Chapter 6.7).

derivative and the pairing between the primal and dual space requires an integral. However, subtleties exist and in infinite dimensions for example the existence of a Hamiltonian function does *not* imply the existence of a Hamiltonian vector field X_H. In practice, one usually formally treats infinite dimensional symplectic manifolds as their finite dimensional counterparts, and verifies required properties in examples when necessary. We will discuss infinite dimensional systems in more detail in Chapter 2.3.4.4.

The space $\mathfrak{X}_{\mathrm{Ham}}(P)$ of Hamiltonian vector fields has an important additional structure which we will discuss next. In Def. 2.97 we introduced the Jacobi-Lie bracket $[\,,]$ and we saw in Chapter 2.3.3 that the space of vector fields $(\mathfrak{X}(P), [\,,])$ equipped with this bracket forms a Lie algebra. For Hamiltonian vector fields we have the following result.

Proposition 2.92. *The space of Hamiltonian vector fields $\mathfrak{X}_{\mathrm{Ham}}(P)$ on a symplectic manifold (P, ω) forms a Lie sub-algebra of the Lie algebra $(\mathfrak{X}(P), [\,,])$ under the Jacobi-Lie bracket $[\,,]$.*

The result in particular implies that $[X_F, X_G] = X_H$ and the Jacobi-Lie bracket of two Hamiltonian vector fields X_F and X_G yields again a Hamiltonian vector field X_H. We will now consider the description of Hamiltonian dynamics using the concept of a Poisson bracket.

Definition 2.212. *Let (P, ω) be a symplectic manifold. The **Poisson bracket** $\{\,,\} : \mathcal{F}(P) \times \mathcal{F}(P) \to \mathcal{F}(P)$ on P is*

$$\{F, G\} = \omega(X_F, X_G)$$

where X_F and X_G are the Hamiltonian vector fields for the Hamiltonian functions $F \in \mathcal{F}(M)$ and $G \in \mathcal{F}(P)$.

Remark 2.156. Poisson brackets can be defined without reference to a symplectic structure and we will refer to such brackets as non-canonical ones. A manifold together with a possibly non-canonical Poisson bracket is a **Poisson manifold** $(P, \{\,,\})$ and these are often considered because they more "robust" in that they apply to a larger class of physical systems. The analogue of a symplectic transformation in the Poisson category is a **Poisson map** $\varphi : P_1 \to P_2$ between two Poisson manifolds $(P_1, \{\,,\})_1)$ and $(P_2, \{\,,\})_2)$ satisfying

$$\varphi^* \{F, G\}_2 = \{\varphi^* F, \varphi^* G\}_1 \tag{2.240a}$$

or equivalently

$$\{F, G\}_1 \circ \varphi = \{F \circ \varphi, G \circ \varphi\}_2 \tag{2.240b}$$

for functions $F, G \in P_2$. On a Poisson manifold, the Hamiltonian vector field can be defined intrinsically as

$$X_H[G] = \mathrm{d}G(X_H) = \{F, G\} \tag{2.241}$$

and its flow $\varphi_t : \mathbb{R} \times P \to P$ is then a Poisson map

$$\varphi_t^* \{F, G\} = \{\varphi_t^* F, \varphi_t^* G\}. \tag{2.242}$$

We will encounter Poisson manifoldss and non-canonical Poisson brackets again when we study systems with symmetries in Chapter 2.3.5.

Example 2.122. Let T^*Q be the cotangent bundle over an n-dimensional configuration manifold Q. From the definition of the canonical symplectic structure on T^*Q it follows that the **canonical Poisson bracket** for functions $F, G : T^*Q \to \mathbb{R}$ is

$$\{F, G\} = \sum_{i=1}^{n} \frac{\partial F}{\partial q^i} \frac{\partial G}{\partial p^i} - \frac{\partial F}{\partial p^i} \frac{\partial G}{\partial q^i}. \tag{2.243}$$

The properties of the Poisson bracket are summarized in the following proposition and they follow mostly from those of the symplectic 2-form.

Proposition 2.93. *Let (P, ω) be a symplectic manifold. Then for function $F, G, H \in \mathcal{F}(P)$ and constants $a, b, c, d \in \mathbb{R}$ the canonical Poisson bracket $\{\,,\}$ on P satisfies:*

i) bilinearity, $\{a\,F + b\,G\,,\,H\} = a\,\{F, H\} + b\,\{G, H\}$ and $\{F\,,\,c\,G + d\,H\} = c\,\{F, G\} + d\,\{F, H\}$;

ii) anti-symmetry, $\{F, G\} = -\{G, F\}$ (or equivalently $\{F, F\} = 0$);

iii) Jacobi identity, $\{F\,,\{G, H\}\} - \{G\,,\{F, H\}\} + \{H\,,\{F, G\}\} = 0$.

When $X_H \in \mathfrak{X}_{\mathrm{Ham}}(P)$ is the Hamiltonian vector field for the Hamiltonian H with flow φ_t we also have

iv) Leibniz rule: $X_H(\{F, G\}) = \{X_H(F), G\} + \{F, X_H(G)\}$;

v) $\varphi_t^* \{F, G\} = \{\varphi_t^* F, \varphi_t^* G\}$;

vi) $\{F, G\} \circ \varphi_t = \{F \circ \varphi_t, G \circ \varphi_t\}$.

Some intuition for the Poisson bracket is obtained with the following identities which relate it to the Lie derivative.

Proposition 2.94. *Let (P, ω) be a symplectic manifold with canonical Poisson bracket $\{\,,\}$. Then for functions $F \in \mathcal{F}(P)$ and $G \in \mathcal{F}(P)$ with Hamiltonian vector fields $X_F \in \mathfrak{X}_{\mathrm{Ham}}(P)$ and $X_G \in \mathfrak{X}_{\mathrm{Ham}}(P)$ it holds*

$$\{F, G\} = X_G[F] = -X_F[G]. \tag{2.244}$$

Proof. From the definition of the Poisson bracket we have

$$\{F, G\} = \omega(X_F, X_G) \tag{2.245a}$$

$$= \mathrm{i}_{X_F} \omega(X_G) \tag{2.245b}$$

and with those for the Hamiltonian vector field one obtains

$$\{F, G\} = \mathrm{d}F(X_G). \tag{2.245c}$$

By the definition of the Lie derivative for functions and Def. 2.96 for the directional derivative the last equation can be written as

$$\{F, G\} = X_G[F] = \pounds_{X_G}F. \tag{2.245d}$$

which proofs the claim, and $\{F, G\} = -X_F(G)$ can be obtained analogously when the anti-symmetry of the Poisson bracket is employed. $\qquad\blacksquare$

By the above proposition, we can think of the Poisson bracket $\{F, G\} = X_G(F)$ as the infinitesimal change in F when "dragged" along the Hamiltonian vector field X_G defined by G, and analogously for $\{F, G\} = -X_F(G)$. The flow φ_t along a Hamiltonian vector field is related to the Poisson bracket as follows.

Proposition 2.95. *Let (P, ω) be a symplectic manifold with canonical Poisson bracket $\{\,,\}$, and let $H \in \mathcal{F}(P)$ be a Hamiltonian with Hamiltonian vector field $X_H \in \mathfrak{X}_{\mathrm{Ham}}(P)$ generating the flow $\varphi_t : \mathbb{R} \times P \to P$. Then*

$$\frac{d}{dt}\left(F \circ \varphi_t\right) = \{F \circ \varphi_t, H\} = \{F, H\} \circ \varphi_t$$

*which is often written more compactly as the **equation of motion in Poisson bracket form***

$$\dot{F} = \{F, H\}.$$

Proof. The Lagrangian derivative dF/dt can be written using the pullback as

$$\frac{d}{dt}\left(F \circ \varphi_t\right) = \frac{d}{dt}\varphi_t^* F \tag{2.247a}$$

and by the definition of the Lie derivative one obtains

$$\frac{d}{dt}\left(F \circ \varphi_t\right) = \varphi_t^*\left(\pounds_{X_H}F\right). \tag{2.247b}$$

With the definition of the pullback and Proposition 2.94 this becomes

$$\frac{d}{dt}\left(F \circ \varphi_t\right) = \{F \circ \varphi_t, H\}. \tag{2.247c}$$

The Hamiltonian H is invariant under the flow so that $H \circ \varphi_t = H$, and hence

$$\frac{d}{dt}\left(F \circ \varphi_t\right) = \{F \circ \varphi_t, H \circ \varphi_t\} \tag{2.247d}$$

and using Proposition 2.93 we obtain

$$\frac{d}{dt}\left(F \circ \varphi_t\right) = \{F, H\} \circ \varphi_t \tag{2.247e}$$

which establishes the result. $\qquad\blacksquare$

The equation of motion in Poisson bracket form describes the time evolution of a function or density on phase space, and it is thus of particular relevance in continuum mechanics. It also plays an important role for systems with symmetry where the canonical Poisson bracket is related to a reduced Lie bracket on the Lie algebra of the symmetry group.

Remark 2.157. Recall from Remark 2.154 that Casimirs C_H are functions other than the Hamiltonian that are preserved along the flow of a Hamiltonian system. With the Poisson bracket these are characterized by

$$0 = \{C_H, H\}. \tag{2.248}$$

The Poisson bracket is an automorphism of the space $\mathcal{F}(P)$ of smooth functions on P. An immediate consequence of its properties in Proposition 2.93 is the following, cf. Def. 2.178.

Corollary 2.20. *Let* (P, ω) *be a symplectic manifold and* $\mathcal{F}(P)$ *the space of smooth functions on* P *with the canonical Poisson bracket* $\{,,\}$. *Then* $(\mathcal{F}(P), \{,\})$ *forms a Lie algebra.*

Remark 2.158. The space $(\mathcal{F}, \{,\})$ is sometimes denoted as the space of observables of a Hamiltonian system.[160] Our discussion on measurements from Chapter 2.3.4.1 hence also applies to dynamics described on phase space with the usual "lifting", for example an observable is a function or density on phase and the Hamiltonian vector field is a dynamics variable.

In the foregoing propositions we established various connections between the Poisson bracket of functions and the associated Hamiltonian vector fields. Intrinsically, this relationship is explained by the following result.

Proposition 2.96. *Let* (P, ω) *be a symlectic manifold. Then the Lie algebra* $(\mathcal{F}(P), \{,\})$ *of smooth functions under the Poisson bracket is homomorphic to the Lie algebra* $(\mathfrak{X}_{\mathrm{Ham}}(P), [,])$ *of Hamiltonian vector fields under the Jacobi-Lie bracket and the homomorphism is given by*

$$[X_F, X_G] = -X_{\{F,G\}}$$

for $F, G \in \mathcal{F}(P)$ *and* $X_F, X_G \mathfrak{X}_{\mathrm{Ham}}(P)$ *the associated Hamiltonian vector fields, respectively.*

Proof. We have to show that $[X_F, X_G] = X_{\{F,G\}}$ for all $F, G \in \mathcal{F}(P)$ with Hamiltonian vector fields $X_F, X_G \in \mathfrak{X}_{\mathrm{Ham}}(P)$. When we consider the vector $[X_F, X_G] \in \mathfrak{X}_{\mathrm{Ham}}(P)$ defined by the Jacobi-Lie bracket as acting on a function $H \in \mathcal{F}(P)$ through the directional derivative, we have by the Leibniz rule

$$[X_F, X_G](H) = X_F X_G(H) - X_G X_F(H) \tag{2.249a}$$

---

[160]The terminology appears in particular in the literature on (geometric) quantization, see for example (Bates and Weinstein, *Lectures on the Geometry of Quantization*).

and by Proposition 2.94 this is equivalent to

$$[X_F, X_G](H) = X_F\{H, G\} - X_G\{H, F\}. \tag{2.249b}$$

Once again using Proposition 2.94 yields

$$[X_F, X_G](H) = \{\{H, G\}, F\} - \{\{H, F\}, G\} \tag{2.249c}$$

and with the Jacobi identity this becomes

$$[X_F, X_G](H) = \{H, \{F, G\}\}. \tag{2.249d}$$

With Proposition 2.94 used a last time we have

$$[X_F, X_G](H) = X_{\{F, G\}}(H) \tag{2.249e}$$

which proves the proposition. $\qquad \blacksquare$

Two additional result that will prove useful in the sequel are the following proposition and corollary. As usual, we will assume that all functions lie in appropriate function space such that the subsequent integrals are well defined.

Proposition 2.97. *Let (P, ω) be a symplectic manifold with boundary ∂P and Liouville form ϖ, and let $F, G \in \mathcal{F}(P)$. Then*

$$\int_P \{F, G\} \varpi = \int_{\partial P} F \, \mathrm{i}_{X_G} \varpi = -\int_{\partial P} G \, \mathrm{i}_{X_F} \varpi.$$

Proof. Using Proposition 2.94 we have

$$\int_P \{F, G\} \varpi = \int_P X_G(F) \varpi \tag{2.250a}$$

and by the invarinace of the Liouville form under the flow and using Proposition 2.76 this can be written as

$$\int_P \{F, G\} \varpi = \int_P \pounds_{F X_G} \varpi. \tag{2.250b}$$

Using Cartan's formula we obtain

$$\int_P \{F, G\} \varpi = \int_P \mathrm{d} \mathrm{i}_{F X_G} \varpi + \mathrm{i}_{F X_G} \mathrm{d} \varpi \tag{2.250c}$$

where the last term vanishes since ϖ is a volume form. With Stokes theorem we hence obtain

$$\int_P \{F, G\} \varpi = \int_{\partial P} \mathrm{i}_{F X_G} \varpi \tag{2.250d}$$

and by Proposition 2.60, ii) we have

$$\int_P \{F, G\} \varpi = \int_{\partial P} F \, \mathrm{i}_{X_G} \varpi \tag{2.250e}$$

which establishes the result. The second part follows from the anti-symmetry of the Poisson bracket. $\qquad \blacksquare$

Corollary 2.21. *Let (P, ω) be a symplectic manifold with boundary ∂P and Liouville form ϖ, and let $F, G, H \in \mathcal{F}(P)$ with at least one of the functions vanishing at the boundary ∂P or having compact support. Then*

$$\int_P F \{G, H\} \, \varpi = \int_P \{F, G\} \, H \, \varpi.$$

Proof. From Proposition 2.97 it follows that

$$0 = \int_P \{HF, G\} \, \varpi \tag{2.251}$$

since at least one of the functions vanishes on the boundary ∂P. But using the Leibniz rule for the Poisson bracket,

$$\{HF, G\} = H \{F, G\} + \{H, G\} F, \tag{2.252}$$

we immediately have

$$0 = \int_P H \{F, G\} \, \varpi + \int_P \{H, G\} \, F \, \varpi \tag{2.253}$$

which implies the result by the anti-symmetry of the Poisson bracket. $\qquad\blacksquare$

2.3.4.4   Infinite Dimensional Hamiltonian Systems[161]

Continuum theories such as electromagnetism, elasticity, fluid dynamics, but also the general theory of relativity are examples of infinite dimensional systems where a configuration is described by a field—a function or tensor (density) in an infinite dimensional function space. In this section we we will briefly discuss infinite dimensional Hamiltonian systems to obtain a "flavour" of their treatment, although development will be less formal than in the foregoing sections. This will prove useful in the next chapter when we study the geometry of light transport which is an infinite dimensional system with symmetry.

As one would expect from the preceding section where we usually did not make a distinction, infinite dimensional Hamiltonian systems can be treated like their finite dimensional counterparts when partial derivatives are replaced by functional derivatives and the usual "dot product" pairing is substituted with an integral. In fact, one could argue that infinite dimensional systems are simpler than their finite dimensional counterparts defined on configuration manifolds since function spaces, which provide the configuration space for continuum theories, are linear spaces. This for example implies that the tangent bundle is trivial, and that each fiber is isomorphic to the base, that is $TV \cong V \times V$ for an infinite dimensional vector space V.[162] In the infinite dimensional case the

---

[161]The material in this section is largely drawn from (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, Chapter 3) and (Schmid, "Infinite-Dimensional Hamiltonian Systems"). The reader should refer there and to (Chernoff and Marsden, *Properties of Infinite Dimensional Hamiltonian Systems*; Marsden and Hughes, *Mathematical Foundations of Elasticity*) for further details.

[162]We will concentrate in thesis on classical infinite dimensional systems. In quantum field theory, for example, one does encounter curved space whose fibers are Hilbert spaces.

difficulties arise from delicate functional analytic question, for example when functionals over infinite function spaces are considered, and it is also in this setting that the the difference between a weakly and strongly non-degenerate pairing arises and becomes important. However, following the literature we will largely disregard functional analytic questions and treat infinite dimensional Hamiltonian systems formally.

As an example for an infinite dimensional symplectic manifold we consider the following.

Example 2.123. Let $\mathcal{F}(\mathbb{R}^n)$ and $\text{Den}(\mathbb{R}^n)$ be defined as usual. Then the L_2 inner product

$$\langle F, g \rangle = \int F g = \int F G \, dx^n \tag{2.254}$$

provides a weakly non-degenerate pairing for all $F \in \mathcal{F}(\mathbb{R}^n)$ and $g = G \, dx^n \in \text{Den}(R^n)$ and

$$\omega\left((F_1, g_1), (F_2, g_2)\right) = \int F_1 g_2 - \int g_1 F_2 \tag{2.255}$$

is the **canonical symplectic form** for the symplectic vector space $\mathcal{F}(\mathbb{R}^n) \times \text{Den}(\mathbb{R}^n)$.

For an infinite dimensional system, Hamilton's principle has the usual form

$$0 = \frac{\delta S}{\delta \varphi(t)} = \frac{d}{d\epsilon} S(\varphi(t) + \epsilon \, \psi(t)) \tag{2.256}$$

for the action functional

$$S\left(\varphi(t)\right) = \int_{t_0}^{t_1} L\left(\varphi(t), \dot{\varphi}(t)\right) dt. \tag{2.257}$$

which depends on a **field Lagrangian** $L(\varphi, \dot{\varphi})$ for a **field** $\varphi(x, t) \in \mathcal{T}_s^r(Q)$, and a "path" $\varphi(t)$ is a function that smoothly varies in time.

Example 2.124. Consider an ideal Euler fluid with velocity field $v \in \mathfrak{X}(Q) \cong \mathcal{T}_0^1(Q)$ flowing in a smooth domain $Q \subset \mathbb{R}^3$. Then the field Lagrangian for the system is

$$L(v(q, t)) = \int_Q \|v(q, t)\|^2 \, dq \tag{2.258}$$

where $\| \cdot \|$ denotes the usual Euclidean norm in $\mathbb{R}^3$.

Example 2.125. Let $\mathcal{A}(Q) = \mathfrak{X}(Q) \cong \mathcal{T}_0^1(Q)$ be the space of magnetic vector potentials on a compact subset $Q \subset \mathbb{R}^3$ such that the magnetic field B is given by $B = -\nabla \times A$ for $A \in \mathcal{A}(Q)$. The Lagrangian for electromagnetic theory is then

$$L(A, \dot{A}) = \frac{1}{2} \int \left(\|\dot{A}\|^2 - \|\nabla \times A\|^2\right) dq \tag{2.259}$$

where $\| \cdot \|$ denotes the usual Euclidean norm in $\mathbb{R}^3$ and $(A, \dot{A}) = T\mathcal{A} = \mathcal{A} \times \mathcal{A}$ since $\mathcal{A}$ is a vector space.

For an infinite dimensional system, the Euler-Lagrange equations that one obtains from Hamilton's action principle are

$$\frac{d}{dt}\frac{\delta L}{\delta \dot{\varphi}} - \frac{\delta L}{\delta \varphi} = 0 \tag{2.260}$$

and these describe the time evolution of the system on the tangent bundle $TV = V \times V$. As in the finite dimensional case, the Hamiltonian description of a system is obtained using the Legendre transform and it "lifts" the system from the tangent bundle $TV \cong V \times V$ to the cotangent bundle $T^*V \cong V^* \times V$ where V^* is the dual space of V. With the **field momentum** $\pi \in V^*$ given by $\pi = \delta L/\delta \dot{\varphi}$, the **field Hamiltonian** is defined as

$$H(\varphi, \pi) = \left(\int \pi \, \dot{\varphi} \right) - L(\varphi, \dot{\varphi}). \tag{2.261}$$

Hamilton's equations governing the time evolution of the field $\varphi \in V$ and its momentum $\pi \in V^*$ then take the form

$$\dot{\varphi} = \frac{\partial \varphi}{\partial t} = \frac{\delta H}{\delta \pi} \tag{2.262a}$$

$$-\dot{\pi} = \frac{\partial \varphi}{\partial t} = \frac{H}{\delta \varphi}. \tag{2.262b}$$

and the Hamiltonian vector field is $X_H = (\dot{\varphi}, \dot{\pi}) = (\delta H/\delta \pi, -\delta H/\delta \varphi) \in T(V^* \times V) \cong (V^* \times V^*) \times (V \to V)$.

Example 2.126. Let us show that Hamilton's equations in Eq. 2.262 hold for the infinite dimensional symplectic vector space in Example 2.123. With the Hamiltonian vector field X_H above and an arbitrary tangent vector $Y = (\delta \psi, \delta \nu) \in T(\mathcal{F}(\mathbb{R}^n) \times \text{Den}(\mathbb{R}^n))$ we have for the symplectic 2-form

$$\omega(X_H, Y) = \omega \left(\left(\frac{\delta H}{\delta \pi}, -\frac{H}{\delta \varphi} \right), (\delta \psi, \delta \nu) \right) \tag{2.263a}$$

and using the concrete expression for ω from Eq. 2.255 this becomes

$$\omega(X_H, Y) = \int \frac{\delta H}{\delta \pi} \, \delta \nu - \int \frac{H}{\delta \varphi} \, \delta \pi. \tag{2.263b}$$

Writing the integral as pairing yields

$$\omega(X_H, Y) = \left\langle \frac{\delta H}{\delta \pi}, \delta \nu \right\rangle - \left\langle \frac{\delta H}{\delta \varphi}, \delta \psi \right\rangle \tag{2.263c}$$

and this can be interpreted as directional derivatives

$$\omega(X_H, Y) = \mathbf{D}_\pi H(\varphi, \pi) \cdot \delta \nu - \mathbf{D}_\varphi H(\varphi, \pi) \cdot \delta \psi. \tag{2.263d}$$

The directional derivatives $\mathbf{D}_\pi$ and $\mathbf{D}_\varphi$ together form the exterior derivative $\mathrm{d}H$ of the Hamiltonian H so that

$$\omega(X_H, Y) = \mathrm{d}H(\varphi, \pi) \cdot (\delta\psi, \delta\nu) \tag{2.263e}$$

which is just the defining equation $\mathrm{i}_{X_H}\omega = \mathrm{d}H$ for the Hamiltonian vector field X_H from Def. 2.211 since $(\delta\psi, \delta\nu)$ was an arbitrary tangent vector.

Example 2.127. As a concrete physical system, consider a classical wave describing the displacement of a field $\varphi \in \mathcal{F}(\mathbb{R}^3)$. The configuration space and resulting symplectic manifold is again those introduced in Example 2.123. In geometrized units, the Hamiltonian for the system is given by

$$H(\varphi, \pi) = \int_Q \left[\frac{1}{2}\pi^2 + \frac{1}{2}\|\nabla\varphi\|^2 + U(\varphi) \right] dq \tag{2.264}$$

where the potential $U(\varphi)$ models nonlinear effects. The Hamiltonian vector field is obtained by computing the functional derivatives in Hamilton's equations. Hence, for the derivative with respect to the field momentum we obtain

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} H(\varphi, \pi + \epsilon\nu) \tag{2.265a}$$

$$= \frac{d}{d\epsilon} \int_Q \left[\frac{1}{2}(\pi + \epsilon\nu)^2 + \frac{1}{2}\|\nabla\varphi\|^2 + U(\varphi) \right] dq \bigg|_{\epsilon=0} \tag{2.265b}$$

$$= \int_Q \left[\frac{d}{d\epsilon} \frac{1}{2}(\pi + \epsilon\nu)^2 \right] dq \bigg|_{\epsilon=0} \tag{2.265c}$$

$$= \int_Q [(\pi + \epsilon\nu)\,\nu]\, dq \bigg|_{\epsilon=0} \tag{2.265d}$$

$$= \int_Q [\pi]\,\nu\, dq \tag{2.265e}$$

so that by the definition of the functional derivative we have

$$\frac{\delta H}{\delta\pi} = \pi. \tag{2.265f}$$

Analogously, for the derivative with respect to the field variable we have

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} H(\varphi + \epsilon\psi, \pi) \tag{2.266a}$$

$$= \frac{d}{d\epsilon} \int_Q \left[\frac{1}{2}(\pi)^2 + \frac{1}{2}\|\nabla(\varphi + \epsilon\psi)\|^2 + U(\varphi + \epsilon\psi) \right] dq \bigg|_{\epsilon=0} \tag{2.266b}$$

$$= \int_Q \left[\frac{d}{d\epsilon} \frac{1}{2}\|\nabla(\varphi + \epsilon\psi)\|^2 + \frac{d}{d\epsilon} U(\varphi + \epsilon\psi) \right] dq \bigg|_{\epsilon=0} \tag{2.266c}$$

$$= \int_Q \left[\frac{\|\nabla(\varphi + \epsilon\,\psi)\|}{\|\nabla(\varphi + \epsilon\,\psi)\|} \frac{d}{d\epsilon} \nabla(\varphi + \epsilon\,\psi) + \frac{\delta U(\varphi + \epsilon\,\psi)}{\delta(\varphi + \epsilon\,\psi)}\,\psi \right] dq \Bigg|_{\epsilon=0}.$$

$$\tag{2.266d}$$

By the definition of the directional derivative, cf. Proposition 2.40, we have $d/d\epsilon\,\nabla(\varphi + \epsilon\,\psi) = -\nabla\cdot\nabla\varphi\cdot\psi$ and we therefore obtain

$$\frac{d}{d\epsilon}\Bigg|_{\epsilon=0} H(\varphi + \epsilon\,\psi, \pi) \tag{2.266e}$$

$$= \int_Q \left[-\nabla\cdot\nabla\varphi\cdot\psi + \frac{\delta U(\varphi)}{\delta\varphi}\,\psi \right] dq \tag{2.266f}$$

$$= \int_Q \left[-\nabla^2\varphi + \frac{\delta U(\varphi)}{\delta\varphi} \right] \psi\,dq \tag{2.266g}$$

so that the functional derivative is given by

$$\frac{\delta H}{\delta\varphi} = -\nabla^2\varphi + \frac{\delta U(\varphi)}{\delta\varphi} \tag{2.266h}$$

The Hamiltonian vector field for the system is therefore

$$X_H = \left(\frac{\partial\varphi}{\partial t}, \frac{\partial\pi}{\partial t} \right) = \left(\frac{\delta H}{\delta\pi}, -\frac{\delta H}{\delta\varphi} \right) = \left(\pi, \nabla^2\varphi - \frac{\delta U(\varphi)}{\delta\varphi} \right) \tag{2.267}$$

and writing it as a second order system we obtain

$$\frac{\partial^2\varphi}{\partial^2 t} = \nabla^2\varphi - \frac{\delta U}{\delta\varphi} \tag{2.268}$$

which recovers the usual inhomogeneous wave equation.

With the previous correspondence between the finite and infinite dimensional case, it comes at no surprise that for a symplectic manifold of the form $V^* \times V$ for some infinite dimensional vector space V the canonical Poisson bracket is given by

$$\{F, G\} = \int \left(\frac{\delta F}{\delta\varphi} \frac{\delta G}{\delta\pi} - \frac{\delta F}{\delta\pi} \frac{\delta G}{\delta\varphi} \right) dq = \left\langle \frac{\delta F}{\delta\varphi}, \frac{\delta G}{\delta\pi} \right\rangle - \left\langle \frac{\delta F}{\delta\pi}, \frac{\delta G}{\delta\varphi} \right\rangle, \tag{2.269}$$

which is the analogue of Eq. 2.243 in the finite dimensional setting.

Example 2.128. In Example 2.125 we introduced the space $\mathcal{A}(Q) \cong \mathcal{T}_0^1(Q)$ of magnetic vector potentials which provides the configuration space for electromagnetic theory. The phase space for the system is hence given by $T^*\mathcal{A} = \mathcal{A}^* \times \mathcal{A}$ where $\mathcal{A}^*$ denotes the dual space of $\mathcal{A}$. The electric field E is defined in the fiber, that is $(A, E) \in T^*\mathcal{A}$, and the magnetic field is $B = -\nabla \times A$, as already discussed. The field Hamiltonian for electromagnetic theory is

$$H(A, E) = \frac{1}{2} \int_Q \left(\|E\|^2 + \|\nabla \times A\|^2 \right) dq \tag{2.270}$$

and the canonical Poisson bracket for the system is hence

$$\{F, G\} = \int \left(\frac{\delta F}{\delta A} \frac{\delta G}{\delta E} - \frac{\delta F}{\delta E} \frac{\delta G}{\delta E} \right) dq. \tag{2.271}$$

The equations describing the time evolution are most conveniently obtained not using the Poisson bracket but with Hamilton's equations. Again computing the necessary functional derivatives we have

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} H(A, E + \epsilon Y) = \left. \frac{d}{d\epsilon} \frac{1}{2} \int_Q \left(\|E + \epsilon Y\|^2 + \|\nabla \times A\|^2 \right) dq \right|_{\epsilon=0} \tag{2.272a}$$

$$= \left. \int_Q \left(\frac{d}{d\epsilon} \frac{1}{2} \|E + \epsilon Y\|^2 \right) dq \right|_{\epsilon=0} \tag{2.272b}$$

$$= \left. \int_Q \left(\frac{d}{d\epsilon} \|E + \epsilon Y\| \frac{E + \epsilon Y}{\|E + \epsilon Y\|} Y \right) dq \right|_{\epsilon=0} \tag{2.272c}$$

$$= \int_Q (E) Y \, dq \tag{2.272d}$$

and analogously for the derivative with respect to the vector potential

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} H(A + \epsilon Z, E) = \left. \frac{d}{d\epsilon} \frac{1}{2} \int_Q \left(\|E\|^2 + \|\nabla \times (A + \epsilon Z)\|^2 \right) dq \right|_{\epsilon=0} \tag{2.273a}$$

$$= \left. \int_Q \left(\frac{d}{d\epsilon} \frac{1}{2} \|\nabla \times (A + \epsilon Z)\|^2 \right) dq \right|_{\epsilon=0} \tag{2.273b}$$

$$= \left. \int_Q \left(\frac{d}{d\epsilon} (\nabla \times (A + \epsilon Z)) \right) dq \right|_{\epsilon=0} \tag{2.273c}$$

and using the definition of the directional derivative as in Example 2.127 we obtain

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} H(A + \epsilon Z, E) = \int_Q (-\nabla \times (\nabla \times A)) Z \, dq. \tag{2.273d}$$

The Hamiltonian vector field for electromagnetism is hence given by

$$X_H = \left(\frac{\partial A}{\partial t}, \frac{\partial E}{\partial t} \right) = \left(\frac{\delta H}{\delta \pi}, -\frac{\delta H}{\delta \varphi} \right) = (E, -\nabla \times (\nabla \times A)) \tag{2.274}$$

and using that $B = -\nabla \times A$ this is equivalent to

$$\frac{\partial E}{\partial t} = \nabla \times B \tag{2.275a}$$

$$\frac{\partial B}{\partial t} = -\nabla \times E \tag{2.275b}$$

which are the two dynamic laws of Maxwell's classical equations. The divergence freeness of the magnetic field follows immediately from its definition as the curl of the vector potential. The fourth of Maxwell's equations, $\nabla \cdot E = \rho$, is related to the gauge symmetry of electromagnetism, and additional structure is required to derive it, cf. Example 2.139.

From the discussion in this section it should be clear that there is a remarkable correspondence between finite and infinite dimensional systems when a geometric formulation of Hamiltonian mechanics is employed. This unification is one of the decisive advantages of geometric mechanics.

2.3.4.5   Contact Mechanics[163]

In this section we will consider contact geometry, the odd dimensional sibling of symplectic geometry. Historically, it has always been in the shadow of its even dimensional counterpart, although many results in the symplectic category have a correspondence for contact manifolds and the two can be considered as duals of each other. In many applications, contact manifolds arise for time varying phenomena with $\mathbb{R} \times \mathcal{M}$ as configuration space where $\mathcal{M}$ is a symplectic manifold as discussed in the previous section.[164] Contact geometry also provides the geometric structure for Hamiltonian systems homogeneous of degree one, which are of importance for light transport, and much of the development in this section is geared towards this applications.

Contact manifolds We will begin by establishing elementary notions of contact geometry.

Definition 2.213. *Let $\mathcal{M}$ be a finite dimensional manifold. A **hyperplane field** ξ **on** $\mathcal{M}$ is a sub-bundle of the tangent bundle $T\mathcal{M}$ of co-dimension one. Locally, any hyperplane field is described by the kernel of a 1-form $\alpha_\xi \in \Omega(\mathcal{M})$.*

Hyperplanes can be described by their normal vector in $T\mathcal{M}$ and these are characterized by their duals in $T^*\mathcal{M}$. Hence, a hyperplane field can indeed be defined by a co-vector or 1-form in $T^*\mathcal{M} \cong \Omega^1(\mathcal{M})$. This also shows that the local 1-form defining a hyperplane field is by no means unique but that there exists an equivalence class of forms defining the same field and which differ by a nonzero constant.

Example 2.129. Let $\mathcal{M} = \mathbb{R}^3$. Then for example any plane which is identical in each tangent space and which from point to point only differs by a translation defines a hyperplane field. For instance, $dz = 0$ defines a hyperplane field

---

[163]The material in this section is largely drawn from the survey article by Etnyre ("Contact Manifolds") and the background sections of (Ratiu and Schmid, "The differentiable structure of three remarkable diffeomorphism groups"; Adams, Ratiu, and Schmid, "A Lie group structure for pseudodifferential operators"; Adams, Ratiu, and Schmid, *A Lie group structure for Fourier integral operators*; Dragulete, Ornea, and Ratiu, *Cosphere bundle reduction in contact geometry*; Dragulete, Ratiu, and Rodriguez-Olmos, "Singular cosphere bundle reduction"). Contact Hamiltonians are discussed in (Geiges, "Contact Geometry").

[164]A point in case is the work by van Hove (*On Certain Unitary Representations of an Infinite Group of Transformations*; "Sur le probléme des relations entre les transformations unitaires de la Mécanique quantique et les transformations canoniques de la Mécanique classique.") where it is shown that not canonical transformations but transformations preserving a 1-form on an odd dimensional manifold $\mathbb{R} \times \mathcal{M}$ are the natural transformations in Hamiltonian mechanics, see also our discussion in Chapter 2.3.5.3.
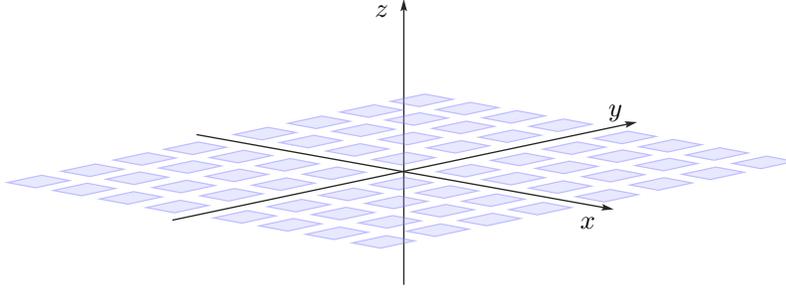
Figure 2.43: Hyperplane field $dz = 0$ in $\mathbb{R}^3$.

formed by planes parallel to the x-y plane, cf. Fig. 2.43. The same hyperplane field is defined by $3dz = 0$ and any scalar multiple $a\,dz = 0$ for $a \in \mathbb{R}^+$.

Definition 2.214. *A **contact manifold** $(\mathcal{M}, \xi)$ of dimension $2n + 1$ is a smooth manifold with a **contact structure**, which is a maximally non-integrable hyperplane field ξ satisfying the Frobenius non-integrability condition*

$$\alpha_\xi \wedge (\mathrm{d}\alpha_\xi)^k = \alpha_\xi \wedge \underbrace{\mathrm{d}\alpha_\xi \wedge \ldots \wedge \mathrm{d}\alpha_\xi}_{n-\text{times}} \neq 0$$

*where $\alpha_\xi \in \Omega^1(\mathcal{M})$ is the **contact** 1-**form** whose kernel locally represents the hyperplane field as $\xi = \ker(\alpha_\xi)$. The 1-form α_ξ is an **exact contact** 1-**form** when $\alpha \wedge (\mathrm{d}\alpha)^n = \mathrm{vol}(M)$ where $\mathrm{vol}(M)$ is a volume form on $\mathcal{M}$.*

Note that the non-integrability condition in the above definition is independent of the contact 1-form α_ξ which is employed to locally represent the hyperplane field ξ.

Remark 2.159. The opposite of the Frobenius non-integrability condition in Def. 2.214 is Frobenius integrability

$$\alpha \wedge (\mathrm{d}\alpha)^k = 0 \tag{2.276}$$

which can also be written as

$$[X, Y] \in \xi \tag{2.277}$$

for all $X, Y \in \xi$ for a subbundle $\xi \subset T\mathcal{M}$ of the tangent bundle, and where $[\,,]$ denotes the Jacobi-Lie bracket. This should be compared to the definition of integrability and integral manifold in Def. 2.98.

Example 2.130. Let $\mathcal{M} = \mathbb{R}^3$. The classical contact structure is then given by $\alpha = dz - y\,dx$. Intuitively, the non-integrability condition means that that for *no* neighborhood in $\mathbb{R}^3$, however small, there exists a smooth surface—an integral manifold—whose tangent planes coincide locally with the planes defined by the contact structure. For example, for the hyperplane field $dz = 0$ in Fig. 2.43
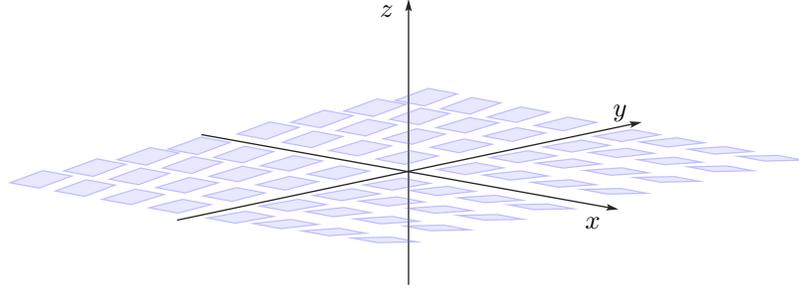
Figure 2.44: The standard contact structure in $\mathbb{R}^3$. The twisting of the planes is such that no smooth 2-surface can be tangent to it in *any* open neighborhood of $\mathbb{R}^3$.

an integral manifold is trivially given by the x-y plane, while in Fig. 2.44 the tangent planes have been twisted such that no integral manifold exists.[165]

Proposition 2.98. *A finite dimensional contact manifold is odd dimensional.*

The result follows immediately from the Frobenius non-integrability condition in Def. 2.214 which can only be satisfied in this case.

Definition 2.215. *Let $(\mathcal{M}, \xi)$ be a contact manifold of dimension $2n + 1$. A **Legendrian manifold** L_ξ is an n-dimensional submanifold of $\mathcal{M}$ such that $TL_\xi \subset \xi$.*

Legendrian manifolds play an important role in the modern theory of partial differential equations and they formed the basis for Lie's work on the subject.[166] They are also intimately related to the Lagrangian submanifolds of a symplectic manifold which were introduced in Def. 2.209.

Canonical or symplectic transformations—diffeomorphisms of a symplectic manifold preserving the symplectic 2-form—played a vital role for the dynamics of Hamiltonian systems. A similar notion for contact manifolds is the following.

Definition 2.216. *Let $(\mathcal{M}_1, \xi_1 = \ker(\alpha_1))$ and $(\mathcal{M}_2, \xi_2 = \ker(\alpha_2))$ be contact manifolds. Then a **contact diffeomorphism** $\varphi_\xi : \mathcal{M}_1 \to \mathcal{M}_2$ from $\mathcal{M}_1$ to $\mathcal{M}_2$ is a diffeomorphism such that $\varphi_\xi^* \alpha_2 = h_\xi \alpha_1$ up to a strictly positive scalar function $h_\xi : \mathcal{M}_1 \to \mathbb{R}$. When h_ξ is the identity section then φ_ξ is an **exact contact diffeomorphism**. When $\mathcal{M}_1 = \mathcal{M}_2$ the infinitesimal generator $X_\xi \in T\mathcal{M}$ of a contact diffeomorphism is a **contact vector field** and conservation of the contact structure is equivalent to*

$$\pounds_{X_\xi} \alpha_\xi = \lambda \, \alpha_\xi$$

for some function $\lambda : \mathcal{M} \to \mathbb{R}^+$.

---

[165]For the intuition of contact structures see also (Arnold, *Mathematical Methods of Classical Mechanics*, Appendix 4.B) and (Spivak, *A Comprehensive Introduction to Differential Geometry*, Chapter 6).

[166]For a modern treatment using this ansatz see (Arnold, *Lectures on Partial Differential Equations*).

An important notion for dynamics on contact manifolds is the following, which provides an infinitesimal characterization of contact transformations.

Definition 2.217. *Let* $(\mathcal{M}, \xi = \ker(\alpha_\xi))$ *be a contact manifold. The* **Reeb vector field** $R \in T\mathcal{M}$ *is the unique vector field satisfying*

$$\mathrm{i}_R\, \alpha_\xi = 1$$

$$\mathrm{i}_R\, \mathrm{d}\alpha_\xi = 0.$$

Example 2.131. Consider again the classical contact structure $\alpha = dz - y\, dx$ for $\mathbb{R}^3$ introduced in Example 2.130. The Reeb vector field is then given by $R_\alpha = \partial/\partial z$. Indeed,

$$\mathrm{i}_{R_\alpha}\alpha = \alpha(R_\alpha) = (dz - y\, dx)\left(\frac{\partial}{\partial z}\right) = dz\left(\frac{\partial}{\partial z}\right) = 1 \qquad (2.278)$$

by the biorthogonality of the basis functions, and

$$\mathrm{i}_{R_\alpha}\, \mathrm{d}\alpha = \mathrm{d}\alpha(R_\alpha, \cdot) = \mathrm{d}\,(dz - y\, dx)\left(\frac{\partial}{\partial z}, \cdot\right) = (dy \wedge dz)\left(\frac{\partial}{\partial z}, \cdot\right) = 0 \quad (2.279)$$

with the result following from $(\alpha \wedge \beta)(v, w) = \alpha(v)\beta(w) - \alpha(w)\beta(v)$.

In the same way as a Hamiltonian vector field preserves the symplectic 2-form along its flow, the Reeb vector field preserves the essential structure of a contact manifold.

Proposition 2.99. *Let* $(\mathcal{M}, \xi = \ker(\alpha_\xi))$ *be a contact manifold with Reeb vector field R. Then R is transverse to the contact hyperplanes and its flow preserves the contact 1-form α_ξ and the contact hyperplane field ξ.*

We saw before that the local structure of a symplectic manifold was characterized by Darboux's theorem, and that locally all symplectic manifolds "look the same". An analogous result for contact manifolds is the following.

Theorem 2.33 (Darboux's theorem for contact manifolds)**.** *All contact manifolds of the same dimension are locally contact diffeomorphic.*

The theorem states that for any two contact manifolds of the same dimensions there exists a contact diffeomorphism in the neighborhood of any two points on the manifolds. Similar to the symplectic case, the interesting structure of contact manifolds is hence global.[167]

Two closely related ways to associate a contact manifold of dimension $2n + 1$ with a symplectic manifold exist: one can "add" a dimension and obtain a symplectic manifold of dimension $2n + 2$, or one can "remove" one dimension to form a symplectic manifold of dimension $2n - 2$. The former approach is

---

[167]Global aspects of contact manifolds are studied in contact topology, see for example (Geiges, *An Introduction to Contact Topology*) for an introduction.

known as symplectification and leads to the contact structure for time varying Hamiltonian systems on a cotangent bundle given by $\alpha = dt + \theta$ where θ is the canonical 1-form on the cotangent bundle. The second approach for obtaining a contact manifold, by "removing" one dimension, will be considered in greater detail in the following, and it will become important later when we consider the geometric structure of light transport.

The contact structure of the cosphere bundle[168]   The cosphere bundle is a contact manifold that can be obtained from a cotangent bundle when its zero section is removed. It arises in various applications in mathematics and physics, and compared to the cotangent bundle it also provides the advantage that it is compact when the base space is compact. We will begin by removing the zero section from the cotangent bundle.

Definition 2.218. *Let Q be a manifold and T^*Q its cotangent bundle. The **slit cotangent bundle** is $T^*Q \backslash \{0\}$.*

The slit cotangent bundle $T^*Q \backslash \{0\}$ is a conic space in each fiber that is closed under multiplication by strictly positive real numbers $a \in \mathbb{R}^+$. The multiplication can be consider as the action of the group $(\mathbb{R}^+, \cdot)$ on $T^*Q \backslash \{0\}$ and it is hence natural to consider the quotient of the slit cotangent bundle by the group, see the foregoing discussion of the action of Lie groups on manifolds in Chapter 2.3.3.2 and also Chapter 2.3.5.

Definition 2.219. *Let Q be a manifold with slit cotangent bundle $T^*Q \backslash \{0\}$. The **cosphere bundle** S^*Q is the quotient*

$$S^*Q = (T^*Q \backslash \{0\}) / \mathbb{R}^+.$$

Remark 2.160. When Q is n-dimensional then the cosphere bundle S^*Q is $(2n-1)$-dimensional.

The cosphere bundle defines an equivalence class or coset $[\alpha_q]$ of covectors $\alpha_q \in T_q^*Q \backslash \{0\}$ in each fiber of the slit cotangent bundle. When one works with S^*Q usually a representative for the cosets is employed, which corresponds to choosing a smooth section in each fiber. A natural choice for the section is available when Q is Riemannian, as the following remark shows.

Remark 2.161. When Q has a Riemannian structure then the tangent bundle TQ and cotangent bundle T^*Q can be identified using the metric. But then also the sphere bundle SQ and the cosphere bundle S^*Q are in correspondence. For the sphere bundle defined as

$$SQ = \{v \in TQ \mid \|v\| = 1\}. \tag{2.280}$$

---

[168]Historically, the contact structure was first studied by Lie ("Zur Theorie Partieller Differentialgleichungen"; *Geometrie der Berührungstransformationen*).

a natural section of the cosphere bundle is given by

$$S^*Q = \left\{ \alpha \in T^*Q \mid \alpha = v^\flat, v \in SQ \right\} \qquad (2.281)$$

where $v^\flat$ is the musical isomorphism which in coordinates is given by $\alpha^j = g_{ij} v^i$, and we omitted inclusion maps. Obviously, the above construction of a section in T^*Q representing the cosphere bundle motivates the nomenclature for the concept.

Remark 2.162. When antipodal points on S_q^*Q are identified, the cosphere bundle coincides with the natural projective space $\mathbb{P}(T_q^*Q\backslash\{0\})$ in each fiber.[169]

The cosphere bundle does not possess a preferred contact structure but an entire family of suitable 1-forms is naturally defined.

Proposition 2.100. *Let $\sigma : S^*Q \to T^*Q\backslash\{0\}$ be a section of the fiber bundle $\pi : T^*Q\backslash\{0\} \to S^*Q$.[170] Then the pullback $\theta_\sigma = \sigma^*\theta$ of the canonical 1-form θ on the slit cotangent bundle $T^*Q\backslash\{0\}$ onto S^*Q is an exact contact 1-form.*

We can hence choose an arbitrary section $\sigma : S^*Q \to T^*Q\backslash\{0\}$ and this will induce a contact structure on the cosphere bundle, and all choices for a section are equivalent.

Remark 2.163. A section $\sigma : S^*Q \to T^*Q\backslash\{0\}$ can be characterized by a strictly positive function $h_\sigma : S^*Q \to \mathbb{R}$ by $\sigma \circ \pi = h_\sigma \, \mathrm{id}_{T^*Q}$ where id_{T^*Q} is the identity section on the slit cotangent bundle.

Remark 2.164. It follows from Remark 2.161 that a preferred contact structure on the cosphere bundle S^*Q exists when Q is Riemannian, and the section $\sigma : S^*Q \to T^*Q\backslash\{0\}$ is then defined by the inclusion map $i : SQ \to TQ$ for the sphere bundle SQ into TQ and Eq. 2.281.

Remark 2.165. When Q is a Riemannian manifold and the cosphere bundle is constructed as in Remark 2.161, then the Reeb vector field R has unit length and it is the generator of the geodesic flow on Q.

The importance of the cosphere bundle and its contact structure arises from the following result which connects it to Hamiltonian mechanics.

Theorem 2.34. *Let a symplectic manifold be given by $(T^*Q\backslash\{0\}, \omega = \mathrm{d}\theta)$, and let $H : T^*Q\backslash\{0\} \to \mathbb{R}$ be a Hamiltonian positive homogeneous of degree one in the fiber variable,*

$$H(q, \lambda p) = \lambda H(q, p)$$

---

[169]In the literature, the cosphere bundle is hence sometimes denoted as oriented projectivized cotangent space.

[170]The bundle projection $\pi : T^*Q\backslash\{0\} \to S^*Q$ is in fact a principal bundle projection since $\mathbb{R}^+$ is acting freely and properly on the conic space $T^*Q\backslash\{0\}$.

*for strictly positive $\lambda \in \mathbb{R}^+$. Then the symplectic flow $\varphi_t : \mathbb{R} \times T^*Q \setminus \{0\} \to T^*Q \setminus \{0\}$ generated by the Hamiltonian vector field X_H preserves the canonical 1-form θ, and the functions homogeneous of degree one form a Lie subalgebra $(\mathcal{F}_0(T^*Q \setminus \{0\}), \{\,\})$ of $(\mathcal{F}(T^*Q \setminus \{0\}), \{\,\})$. Moreover, the symplectic flow φ_t projects onto a contact diffeomorphism $\tilde{\varphi}_t : \mathbb{R} \times S^*Q \to S^*Q$ on the cosphere bundle $(S^*Q, \xi = \ker(\alpha_{xi}))$, and this diffeomorphism commutes with the multiplications action $m_\lambda(q, p) = (q, \lambda p)$ of $\mathbb{R}^+$. The contact vector field X_ξ associated with $\tilde{\varphi}_t$ is characterized by a **contact Hamiltonian** $H_\xi : S^*Q \to \mathbb{R}$ by*

i) $X_\xi \mapsto H_\xi = \alpha_\xi(X_\xi)$

ii) $H \mapsto X_\xi : \mathrm{i}_{X_\xi} \mathrm{d}\alpha = \mathrm{d}H(R)\, \alpha_\xi - \mathrm{d}H$

where R is the Reeb vector field for the contact manifold $(S^*Q, \xi = \ker(\alpha_{xi}))$.

A Hamiltonian system on the slit cotangent bundle with Hamiltonian homogeneous of degree one is hence equivalent to a system on the cosphere bundle, where the dynamics are described with one less variable. A corollary of the above result is that when Q has a Riemannian structure and the cosphere bundle is defined using the Riemannian metric, then the Reeb vector field can be identified with the infinitesimal generator X_H of the Hamiltonian flow.

Remark 2.166. Theorem 2.34 asserts that the symplectic flow on the slit cotangent bundle $T^*Q \setminus \{0\}$ generated by a Hamiltonian homogeneous of degree one does not only preserve the symplectic 2-form $\omega = \mathrm{d}\theta$ but also the canonical 1-form θ. The set of such symplectic transformation forms a subgroup $\mathrm{Diff}^\theta_{\mathrm{can}}(T^*Q \setminus \{0\})$ of the group of all canonical diffeomorphisms $\mathrm{Diff}_{\mathrm{can}}(T^*Q \setminus \{0\})$, and by the above theorem $\mathrm{Diff}^\theta_{\mathrm{can}}(T^*Q \setminus \{0\})$ is as a Lie group isomorphic to the group of contact transformations $\mathrm{Diff}_{\mathrm{con}}(S^*Q)$ on the cosphere bundle S^*Q.[171] For the structure of the group of diffeomorphisms preserving the canonical 1-form the difference between the cotangent bundle T^*Q and the slit cotangent bundle $T^*Q \setminus \{0\}$ is thereby critical: On T^*Q the conservation of the canonical 1-form θ implies that the phase space flow is the cotangent lift of a configuration space diffeomorphism, cf. Example 2.119, which conversely implies that the dynamics on the cotangent bundle can be trivialized to a flow on configuration space Q. On the slit cotangent bundle $T^*Q \setminus \{0\}$ no such trivialization is possible, and $\mathrm{Diff}^\theta_{\mathrm{can}}(T^*Q \setminus \{0\})$ is hence much "larger" than $\mathrm{Diff}^\theta_{\mathrm{can}}(T^*Q)$. Intuitively, the difference between $T^*Q \setminus \{0\}$ and T^*Q arises because the zero section of the cotangent bundle is rigid part under the transformations of interest: a smooth diffeomorphism preserves the canonical 1-form θ if the pairing $\theta(X)$ with a vector X is preserved under the flow which can be satisfied by consistently transforming the 1-form and the vector. However, the pairing of the zero section and a vector is preserved if and

---

[171]The isomorphism between $\mathrm{Diff}^\theta_{\mathrm{can}}(T^*Q \setminus \{0\})$ and $\mathrm{Diff}_{\mathrm{con}}(S^*Q)$ was studied extensively in the context of pseudodifferential and Fourier integral operators, see (Ratiu and Schmid, "The differentiable structure of three remarkable diffeomorphism groups"; Adams, Ratiu, and Schmid, "A Lie group structure for pseudodifferential operators"; Adams, Ratiu, and Schmid, *A Lie group structure for Fourier integral operators*).

only if the zero section is mapped to the zero section, up to a set of measure zero corresponding to the 1-forms being "orthogonal" to the vector.

The equivalence of the dynamics on $T^*Q \setminus \{0\}$ and S^*Q established in Theorem 2.34 is also shown by the fact that the Legendrian submanifolds L_ξ of the contact manifold S^*Q are lifted to Lagrangian submanifold L_ω on T^*Q, and the contact diffeomorphism induced by a Hamiltonian homogeneous of degree one on the cosphere bundle preserves Legendrian submanifolds.[172]

2.3.5 Symmetries and Dynamics on Lie Groups[173]

In this section we will extend the discussion of the previous section to respect and exploit the symmetries that exist in almost all mechanical systems. In Chapter 2.3.5.1, we will formally define what a symmetry is, introduce the important concept of a momentum map, and outline reduction for symplectic manifolds. In Chapter 2.3.5.2, we will discuss in detail Lie-Poisson reduction for systems whose configuration space is a Lie group, and there also the example of an ideal Euler fluid will be thoroughly studied. In Chapter 2.3.5.3, we will investigate the structure of mechanical systems whose configuration space and symmetry group is the group $\mathrm{Diff}_{\mathrm{can}}(P)$ of canonical diffeomorphisms. Throughout, the reader should keep our intuitive discussion of symmetries in Chapter 2.3.1 in mind.

2.3.5.1 Symmetries and Reduction

We will begin by outlining the importance of symmetries in modern physics and geometric mechanics and providing an overview over reduction theory, including an introduction to the concept of a momentum map which is central to the Hamiltonian description of symmetries and which provides a modern geometric formulation of Noether's theorem.

---

[172]Arnold, Gusein-Zade, and Varchenko, *Singularities of Differentiable Maps: The classification of critical points, caustics and wave fronts*, p. 317.

[173]Principal references for the material in this section are (Marsden et al., "Hamiltonian Systems with Symmetry, Coadjoint Orbits and Plasma Physics"; Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*; Holm, Schmah, and Stoica, *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*) and an alternative introduction is available for example in (Marsden and Ratiu, *Mechanical Systems: Symmetry and Reduction*). Our discussion will focus on Hamiltonian systems with symmetry, which was also historically the part of the theory which developed first, although today also reduction for Lagrangian systems is well understood, see for example (Cendra et al., "Lagrangian Reduction, the Euler-Poincaré Equations and Semidirect Products"; Holm, Marsden, and Ratiu, *The Euler-Poincaré Equations and Semidirect Products with Applications to Continuum Theories*; Cendra, Marsden, and Ratiu, *Lagrangian Reduction by Stages*; Cendra, Marsden, and Ratiu, "Geometric Mechanics, Lagrangian Reduction, and Nonholonomic Systems") and the ideas were first developed in (Marsden and Scheurle, "Lagrangian reduction and the double spherical pendulum"; Marsden and Scheurle, "The Reduced Euler-Lagrange Equations").

Symmetries and Reduction   Continuous symmetries are central to most modern physical theories and it would not be exaggerated to claim that contemporary physics is the quest for finding ever more symmetry. For example, Einstein's theory of special relativity is entrenched in the notion of Lorentz invariance, and the standard model, the most sophisticated "theory of everything" conceived so far, is almost completely characterized by a postulated set of symmetries believed to be satisfied by nature. Yet, also for the Euler top, seemingly one of the simplest mechanical systems, the dynamics are firmly rooted in the invariance of the system under the rotation group SO(3).

As was shown by Noether almost a century ago,[174] associated with every continuous symmetry is a conserved quantity, and her theorem paved the way for the paradigm shift towards symmetries in modern physics.[175] Reduction is central when systems with symmetry are studied, and intuitively it can be understood as the simplification of a system's description by exploiting the conserved quantities that are invariant in time.

Group Actions and Symmetries   Mathematically, symmetries of a system are described by Lie groups and their action on the configuration space of a system. We will hence in the following briefly recall some essential notions on Lie group actions from Chapter 2.3.3.2 and connect them with the physical applications of this section.

Recall from Chapter 2.3.3.2 that the left action of a Lie group G on a manifold $\mathcal{M}$ was the smooth mapping $\psi : G \times \mathcal{M} \to \mathcal{M}$ satisfying $\psi_e \circ x = x$ and $\psi_g \circ \psi_h \circ x = \psi_{g \circ h} \circ x$ for all $x \in \mathcal{M}$ and $g \in G$, and a right action was defined analogously satisfying $\psi_g \circ \psi_h \circ x = \psi_{h \circ g} \circ x$. The symmetry of a tensor under a group action can then be defined as follows.

Definition 2.220. *Let G be a Lie group and $\mathcal{M}$ be a manifold, and let $\varphi : G \times \mathcal{M} \to \mathcal{M}$ be an action of G on $\mathcal{M}$. A **tensor field** $t \in \mathcal{T}_s^r(\mathcal{M})$ **is invariant under the action of** G when*

$$\varphi_g^* t = t$$

*for all $g \in G$, and G is then known as the **symmetry group** or **Lie group symmetry** of t.*

The most important examples of tensors whose invariance we will consider are, functions, vector fields, and differential forms.

Remark 2.167. In Def. 2.159, we introduced before the notion of invariance of a differential form under the flow of a vector field. It is easy to see that the two definitions are consistent and Def. 2.159 represents an $(\mathbb{R}, +)$ symmetry under time translation.

[174](Noether, "Invariante Variationsprobleme"), see (Noether and Tavel, *Invariant Variation Problems*) for an English translation.

[175]Poincaré knew the technical results of the special theory of relativity before Einstein, but it was Einstein who understood their significance because for him the results arose from a new symmetry, cf. (Brading and Castellani, "Symmetry and Symmetry Breaking").

As the following two examples show, not alone the group but also the definition of its action is critical for the symmetries of a system.

Example 2.132. Let $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$ and let the action of $(\mathbb{R}, +)$ on $\mathbb{R}^2$ be given by $\psi_z : (x,y) \mapsto (x, y+z)$. Then every function which is constant in the y coordinate is invariant under the group action.

Example 2.133. For any Hamiltonian system (P, ω, H) the flow $\varphi_t : \mathbb{R} \times P \to P$ of the Hamiltonian vector field X_H preserves the symplectic 2-form ω, cf. Proposition 2.90. Hence, ω is invariant under the vector group $(\mathbb{R}, +)$ whose action is given by the Hamiltonian flow φ_t, and it is here again useful to think about the action of $(\mathbb{R}, +)$ as a group homomorphism into $\mathrm{Diff}(\mathcal{M})$, cf. Remark 2.142.

In Chapter 2.3.3.2 we also introduced the infinitesimal generator of a group action φ on $\mathcal{M}$ which was the vector field $\xi_\mathcal{M}$ whose flow is φ, that is

$$\xi_\mathcal{M}(x) = \left. \frac{d}{dt} \right|_{t=0} \left(\varphi_t^\xi \circ x \right) \in T_x \mathcal{M} \tag{2.282}$$

where φ_t^ξ is the one parameter group $\varphi_t^\xi = \exp(t\,\xi)$ generated by the Lie algebra element $\xi \in \mathfrak{g}$. The infinitesimal generator led to the notion of the action $\phi : \mathfrak{g} \times \mathcal{M} \to T\mathcal{M}$ of a Lie algebra $\mathfrak{g}$ on a manifold. We can hence define an infinitesimal analogue of Def. 2.220.

Definition 2.221. *Let $\mathfrak{g}$ be a Lie algebra and $\mathcal{M}$ be a manifold, and let $\phi : \mathfrak{g} \times \mathcal{M} \to T\mathcal{M}$ be an action of $\mathfrak{g}$ on $\mathcal{M}$ given by the infinitesimal generator $\xi_\mathcal{M} \in T\mathcal{M}$ for $\xi \in \mathfrak{g}$. A **tensor field** $t \in \mathcal{T}_s^r(\mathcal{M})$ **is invariant under the action of** $\mathfrak{g}$ when*

$$\pounds_{\xi_\mathcal{M}} t = 0$$

*for all $\xi \in \mathfrak{g}$, and $\mathfrak{g}$ is then known as the **symmetry algebra** or **Lie algebra symmetry** of t.*

In the special but important case of a function $f \in \mathcal{F}(\mathcal{M})$, invariance under a Lie algebra action is hence given by

$$\xi_\mathcal{M}[f] = 0.$$

An example for a group action that infinitesimally preserves a tensor is the following.

Example 2.134. Let $Q = \mathbb{R}^3$ and consider the usual action of $SO(3)$. The cotangent lift of the action is

$$R \circ (q, p) \mapsto (R\,q, R\,p) \tag{2.283}$$

for $(q, p) \in T^* \mathbb{R}^3 \cong \mathbb{R}^3 \times \mathbb{R}^3$ and $R \in SO(3)$, cf. Def. 2.196. With R of the form $R(t) = \exp(t\,\xi)$ and $\xi \in \mathfrak{so}(3)$ a skew symmetric matrix in the Lie algebra

$\mathfrak{so}(3)$ of SO(3), cf. Example 2.93, differentiating the definition of the action with respect to time yields for the infinitesimal generator

$$\xi_P = (\xi\, q, \xi\, p) \ \in T(T^*Q). \tag{2.284}$$

A lengthy but straight-forward calculation also shows that

$$\pounds_{\xi_P}\omega = \mathrm{di}_{\xi_P}\omega + \mathrm{i}_{\xi_P}\mathrm{d}\omega = \mathrm{di}_{\xi_P}\omega = 0 \tag{2.285}$$

and the canonical symplectic 2-form $\omega = -\mathrm{d}\theta$ on T^*Q is preserved under the action of the Lie algebra $\mathfrak{so}(3)$ and of the group SO(3).

The conservation of the symplectic 2-form in the above action of SO(3) on $T^*\mathbb{R}^3$ holds for any cotangent lifted action, which is a corollary to the result that the flow of a cotangent lift of a configuration space diffeomorphism preserves the canonical 1-form on T^*Q, cf. Example 2.119. More generally, we specialize Def. 2.220 and Def. 2.221 as follows for actions preserving the symplectic 2-form.

Definition 2.222. *Let G be a Lie group acting on the symplectic manifold (P, ω) by $\varphi : G \times P \to P$. Then the action is **canonical** or **symplectic** when $\varphi_g^*\omega = \omega$. The action $\mathfrak{g} \times \mathcal{M} \to T\mathcal{M}$ of a Lie algebra is $\mathfrak{g}$ by the infinitesimal generator ξ_P is canonical when $\pounds_{\xi_P}\omega = 0$.*

That the above definitions of a canonical Lie group and Lie algebra action are consistent, and that the latter one implies the former, follows from Proposition 2.90, and we summarize the situation in the following proposition.

Proposition 2.101. *Let $(G, \mathfrak{g})$ be a Lie group acting canonically on a symplectic manifold (P, ω) by $\varphi : G \times P \to P$, and let H be invariant under the group action so that $H \circ \varphi_g = H$ for all $g \in G$. Then H is also infinitesimally invariant under the associated Lie algebra action $\varphi : \mathfrak{g} \times P \to TP$.*

As Example 2.133 showed, for every Hamiltonian system the action of $(\mathbb{R}, +)$ by the flow $\varphi_t : \mathbb{R} \times P \to P$ of a Hamiltonian vector field X_H is canonical, and it is infinitesimally invariant since $\pounds_{X_H}\omega = 0$.

In Def. 2.187 in Chapter 2.3.3.2 we characterized actions of Lie groups, and one important result was Proposition 2.85 which showed that when the action $\varphi : G \times \mathcal{M} \to \mathcal{M}$ of a Lie group G on a manifold $\mathcal{M}$ is free and proper then the quotient space $G/\mathcal{M}$ is a smooth manifold and $\pi : \mathcal{M} \to \mathcal{M}/G$ is a smooth submersion. The result is of particular importance for geometric mechanics since it determines when the reduced phase space of a mechanical system with symmetry is a manifold, and as the following example shows we already encountered a similar situation in the foregoing.

Example 2.135.[176] In Chapter 2.3.4.5 we introduced the cosphere bundle which was the quotient

$$S^*Q = (T^*Q \setminus \{0\}) / \mathbb{R}^+ \tag{2.286}$$

---

[176]We thank Alex Castro for help with this example.
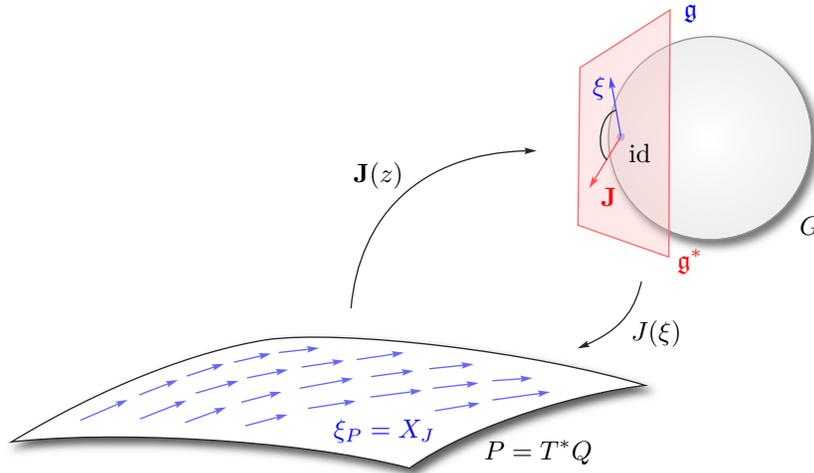
Figure 2.45: The momentum map $\mathbf{J}(z)$ establishes a correspondence between phase space $P = T^*Q$ and the dual Lie algebra $\mathfrak{g}^*$ of the symmetry group, so that the infinitesimal generator ξ_P of the group action coincides with the Hamiltonian vector field X_J for a Hamiltonian $J(\xi) : \mathfrak{g} \to \mathcal{F}(T^*Q)$ given by $J(\xi)(z) = \langle \mathbf{J}(z), \xi \rangle$.

of the action $(q, p) \mapsto (q, \lambda p)$ of the multiplicative group $(\mathbb{R}^+, \cdot)$ on the conic spaces $T_q^*Q \backslash \{0\}$. There, we tacitly assumed that S^*Q is a smooth manifold, which is rather apparent at least when Q is Riemannian, cf. Remark 2.161, but it also follows from the properties of the $(\mathbb{R}^+, \cdot)$ action which is free, since it does not have fixed points, and proper, since for any $S \subset T^*Q \backslash \{0\}$ its pre-image under $(q, p) \mapsto (q, \lambda p)$ for $\lambda \in \mathbb{R}^+$ is clearly compact.

After recapitulating the elementary notions of group actions on manifolds and after defining what we mean by the symmetry of a group action, we are prepared to study how Lie group actions can describe mechanical systems, and this will lead naturally to the notion of a momentum map which is central to the description of symmetries in Hamiltonian mechanics.

Momentum Maps The "glue" between Lie groups and mechanics is provided by the description of a group action as a Hamiltonian system. To see how this is possible consider a canonical action $\varphi : G \times T^*Q \to T^*Q$ of a Lie group G on the cotangent bundle $P = T^*Q$ of a finite configuration manifold Q. When we assume that the action can also be described as a Hamiltonian system, which locally but not globally is guaranteed by Proposition 2.90, then there has to exist a globally Hamiltonian vector field $X_J \in T(T^*Q)$ for some Hamiltonian $J : T^*Q \to \mathbb{R}$, and since the Hamiltonian flow is to coincide with the flow generated by the group, X_J has to coincide with the infinitesimal generate $\xi_P \in T(T^*Q)$ of the element $\xi \in \mathfrak{g}$ in the Lie algebra $\mathfrak{g}$ of G generating the flow. Put differently, to obtain a globally Hamiltonian description we have to assume that for every $\xi \in \mathfrak{g}$ there exists a Hamiltonian function $J_\xi(z) \in \mathcal{F}(T^*Q)$ such

that

$$\xi_P(z) = X_J(z) \tag{2.287}$$

for all $z \in P = T^*Q$. But then $J_\xi(z)$ can also be considered as the map $J : \mathfrak{g} \to \mathcal{F}(T^*Q)$ whose image in $\mathcal{F}(T^*Q)$ for every $\xi \in \mathfrak{g}$ is the sought Hamiltonian such that $\xi_P(z) = X_J(z)$, see Fig. 2.45. When we require $J_\xi(z)$ to be linear in the Lie algebra element $\xi \in \mathfrak{g}$, then we can define the map $J : \mathfrak{g} \to \mathcal{F}(T^*Q)$ also using a linear functional $\mathbf{J}(z)$ on $\mathfrak{g}$ such that $\mathbf{J}_z(\xi) = J_\xi(z)$ and where $\mathbf{J}$ has to depend on the phase space location since ξ. But since $\mathfrak{g}$ is a vector space the space of linear functionals is the dual Lie algebra $\mathfrak{g}^*$ and with $\mathbf{J}(z) \in \mathfrak{g}^*$, the Hamiltonian J can be written as

$$J_\xi(z) = \langle \mathbf{J}(z), \xi \rangle \tag{2.288}$$

where $\langle , \rangle$ is the canonical pairing between $\mathfrak{g}$ and $\mathfrak{g}^*$. But then the linear functional $\mathbf{J}(z)$ can be as a mapping $\mathbf{J} : T^*Q \to \mathfrak{g}^*$ from phase space T^*Q to the dual Lie algebra, and for each $z \in T^*Q$ it assigns an element in $\mathfrak{g}^*$ such that the infinitesimal generator $\xi_\mathcal{M}$ of $\xi \in \mathfrak{g}$ coincides with the Hamiltonian vector field for the Hamiltonian $J_\xi(z) = \langle \mathbf{J}(z), \xi \rangle$, see again Fig. 2.45. The map $\mathbf{J} : T^*Q \to \mathfrak{g}^*$ is known as momentum map,[177] and the name arises since its image in $\mathfrak{g}^*$ can often be identified with momentum which is conserved in many classical systems We are now prepared to more formally describe how a canonical Lie group action can be considered as the flow of a Hamiltonian system.

Definition 2.223. *Let the canonical action $\mathfrak{g} \times P \to TP$ of a Lie algebra $\mathfrak{g}$ on a symplectic manifold (P, ω) be given by the infinitesimal generator ξ_P. Assuming there exists a linear* **momentum Hamiltonian** *$J : \mathfrak{g} \to P$ such that*

$$\xi_P = X_{J(\xi)}$$

then the **momentum map** *$\mathbf{J} : P \to \mathfrak{g}^*$ is the map satisfying*

$$\langle \mathbf{J}(z), \xi \rangle = J(\xi)(z)$$

for all $z \in P$ and $\xi \in \mathfrak{g}$.

Remark 2.168. In the above definition we employed a Lie algebra action instead of a Lie group action since the momentum map only depends on $\xi \in \mathfrak{g}$. The two are related by the exponential map $\exp(t\,\xi) : \mathbb{R} \times \mathfrak{g} \to G$ and for every Lie algebra action one can obtain an associated Lie group action by exponentiation.

Before we consider examples of momentum maps let us state the theorem from which is central to their importance in applications.

---

[177]In the literature, sometimes 'moment map' is used instead of 'momentum map'.

Theorem 2.35 (Hamiltonian version of Noether's theorem). *Let the canonical action* $\mathfrak{g} \times P \to TP$ *of a Lie algebra* $\mathfrak{g}$ *on a symplectic manifold* (P, ω) *be given by the infinitesimal generator* ξ_P, *and let* $\mathbf{J} : P \to \mathfrak{g}^*$ *be a momentum map for the action. Then for a* $\mathfrak{g}$-*invariant Hamiltonian* $H \in \mathcal{F}(P)$ *such that* $\xi_P[H] = 0$ *the flow* $\varphi_t : \mathbb{R} \times P \to P$ *of the Hamiltonian vector field* X_H *conserves the momentum map such that*

$$\mathbf{J} \circ \varphi_t = \mathbf{J}.$$

Proof. The invariance of H under $\mathfrak{g}$ is defined as $\xi_P[H] = 0$. With Proposition 2.94 this is equivalent to $\{J(\xi), H\} = 0$ where $J(\xi)$ is the momentum Hamiltonian for the momentum map $\mathbf{J}$. Hence $J(\xi)$ is a Casimir of the Hamiltonian flow X_H for each $\xi \in \mathfrak{g}$. But when $J(\xi)$ is invariant under the flow then by Def. 2.223 also the image of the momentum map in $\mathfrak{g}^*$ has to be invariant, which proves our claim. $\square$

Next, we will consider some examples for momentum maps where the conserved quantities are well known.

Example 2.136. Continuing Example 2.133, let (P, ω) be again a symplectic manifold and $H : P \to \mathbb{R}$ a time invariant Hamiltonian with Hamiltonian vector field X_H. As we showed before, the flow

$$\varphi_t = \exp(t \, X_H) \tag{2.289}$$

generated by X_H defines an action

$$\varphi_t : \mathbb{R} \times P \to P : \varphi_t(z) = z(t) \tag{2.290}$$

of $\mathbb{R}$ on P, and the action is by construction canonical. For the vector Lie group $(\mathbb{R}, +)$ we also have $\mathfrak{g} = \mathbb{R}$ and $\mathfrak{g}^* = \mathbb{R}$, and the pairing $\langle\,,\rangle$ between the primary and dual Lie algebra is given by scalar multiplication. With Eq. 2.289 and the definition of the infinitesimal generator, we have

$$\xi_P(z) = \frac{d}{dt} \varphi_t(z) = \frac{d}{dt} \exp(t \, X_H) = t \, X_H, \tag{2.291}$$

cf. Def. 2.191. The momentum Hamiltonian $J : \mathfrak{g} \to \mathcal{F}(P) : \mathbb{R} \to \mathcal{F}(P)$ is hence

$$J(t) = tH \tag{2.292}$$

since the Hamiltonian vector field is only determined by the spatial "gradient" of the Hamiltonian H. By definition, the momentum map has to satisfy

$$\langle \mathbf{J}, t \rangle = J(t) = tH. \tag{2.293}$$

But this immediately implies

$$\mathbf{J} = H \tag{2.294}$$

and we can hence interpret the conservation of the Hamiltonian H along the flow of the Hamiltonian vector field X_H as a consequence of Noether's theorem.

Example 2.137. Consider a particle in $\mathbb{R}^3$ as in Table 2.1 and as was our running example in Chapter 2.3.4.2, and let $(\mathbb{R}^3, +)$ act on the system by

$$\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \ : \ q \mapsto q + x \tag{2.295}$$

where $q \in \mathbb{R}^3$ is the particle position and $x \in \mathbb{R}^3$ the translation. The cotangent lift of the action is

$$\mathbb{R}^3 \times T^*\mathbb{R}^3 \to T^*\mathbb{R}^3 \ : \ (q, p) \mapsto (q + x, p) \tag{2.296}$$

where p is the momentum of the particle. With the translation x of the form $x = \exp(t\xi) = t\xi$, where ξ is an element in the Lie algebra $\mathfrak{g} \cong \mathbb{R}^3$, differentiating Eq. 2.296 with respect to t shows that the infinitesimal generator for the lifted action is given by

$$\xi_P = \frac{d}{dt}(q + x, p) = \frac{d}{dt}(q + (t\xi), p) = (\xi, 0). \tag{2.297}$$

By the construction of the momentum map we have to have $\xi_P = X_{J(\xi)}$ for some momentum Hamiltonian $J : g \to \mathcal{F}(T^*Q)$, and since our phase space is a cotangent bundle the Hamiltonian vector field $X_{J(\xi)}$ also has to satisfy Hamilton's equations so that

$$\xi_P = (\xi, 0) = \left(\frac{\partial J(\xi)}{\partial p}, -\frac{\partial J(\xi)}{\partial q} \right) = X_{J(\xi)}. \tag{2.298}$$

From the above equation we immediately see that the momentum Hamiltonian J has to be constant in or independent of q. Eq. 2.298 can be solved for the momentum Hamiltonian by inspection since it is easily seen that the left equation is obtained when J is linear in the momentum. Hence,

$$J(\xi) = p \cdot \xi = \langle p, \xi \rangle \tag{2.299}$$

Comparing to the definition of the momentum map in Def. 2.223,

$$\langle \mathbf{J}(q, p), \xi \rangle = J(\xi)(q, p), \tag{2.300}$$

we immediately see that

$$\mathbf{J}(q, p) = p \ \in \mathfrak{g}^* \cong \mathbb{R}^3 \tag{2.301}$$

and the momentum map is given by the particle's linear momentum as a value in the Lie algebra $\mathfrak{g}^* \cong \mathbb{R}^3$. The example generalizes to an n particle system and the conserved quantity is then the total linear momentum of the system

$$\mathbf{J}(q_i, p^i) = \sum_{i=1}^{n} p^i \tag{2.302}$$

where $p^i = (p_1^i, p_2^i, p_3^i)$ is the linear momentum of the ith particle, and its conservation for a translation invariant system is well known.

Example 2.138. Consider again a particle as in the previous example but let this time the rotation group SO(3) act on the particle's configuration, that is we consider the action

$$\text{SO}(3) \times \mathbb{R}^3 \to \mathbb{R}^3 \ : \ q \mapsto R\,q \tag{2.303}$$

for $R \in \text{SO}(3)$. The cotangent lift of the action is given by

$$\text{SO}(3) \times T^*\mathbb{R}^3 \to T^*\mathbb{R}^3 \ : \ (q,p) \mapsto (R\,q, R\,p). \tag{2.304}$$

Again differentiating the above equation with respect to t for $R = \exp(t\,\xi)$ we obtain for the infinitesimal generator ξ_P of the lifted action

$$\xi_P = (\xi\,q, \xi\,p) \tag{2.305}$$

where $\xi \in \mathfrak{so}(3)$ is an element in the Lie algebra $\mathfrak{so}(3)$ of the rotation group SO(3), and we represent it for the moment using a skew-symmetric matrix. We seek a Hamiltonian vector field X_J such that $\xi_P = X_J$, and since we are again on a cotangent bundle X_J has to satisfy Hamilton's equations. Hence,

$$\xi_P = (\xi\,q, \xi\,p) = \left(\frac{\partial J(\xi)}{\partial p}, -\frac{\partial J(\xi)}{\partial q} \right) = X_J. \tag{2.306}$$

The momentum Hamiltonian J has thus to be linear in both position and momentum which is satisfied when

$$J(\xi) = (\xi\,q) \cdot p = -(\xi\,p) \cdot q. \tag{2.307}$$

Employing the hat map representation for $\mathfrak{so}(3)$ introduced in Example 2.93 so that $\xi \mapsto \check{\xi}$, and using standard identities for the cross product, we have

$$J(\xi) = (\xi\,q) \cdot p = (\check{\xi} \times q) \cdot p = (q \times p) \cdot \check{\xi}. \tag{2.308}$$

Comparing to the definition of the momentum map in Def. 2.223 we obtain

$$\mathbf{J}(q,p) = q \times p \ \in \mathfrak{so}(3) \cong \mathbb{R}^3 \tag{2.309}$$

which is the usual expression for the angular momentum of a particle, and whose conservation for a rotation invariant system is well known.

Example 2.139. In this example we will return to electromagnetic theory which we considered before in Example 2.128. Recall that the configuration space of the system is the space $\mathcal{A}(Q) \cong \mathcal{T}_0^1(Q)$ of magnetic vector potentials $A \in \mathcal{A}(Q)$, and that an element in the cotangent bundle is $(A, -E) \in T^*\mathcal{A}(Q)$ where E is the electric field. Electromagnetic theory is invariant under the gauge group $G = \mathcal{F}(\mathbb{R}^3)$ whose action on $T^*\mathcal{A}$ is

$$\varphi \circ (A, -E) = (A + \nabla\varphi, -E) \tag{2.310}$$

for $\varphi \in \mathcal{F}(\mathbb{R}^3)$. Differentiating the above definition of the action with respect to time for $\varphi = \exp(t\,\xi)$ yields for the infinitesimal generator

$$\xi_P = (\nabla\xi, 0) \tag{2.311}$$

where $\xi \in \mathfrak{g}^* \cong \mathcal{F}(\mathbb{R}^3)$ and the Lie algebra is isomorphic to G since the group is a vector space. The infinitesimal generator has to coincide with a Hamiltonian vector field $X_{J(\xi)}$ and since $T^*\mathcal{A}$ is a cotangent bundle $X_{J(\xi)}$ has to satisfy the infinite dimensional Hamilton's equations, that is we have to have

$$\xi_P = (\nabla\xi, 0) = \left(\frac{\delta J(\xi)}{\delta E}, -\frac{\delta J(\xi)}{\delta A}\right) = X_{J(\xi)}. \tag{2.312}$$

From the above equation we see that the momentum Hamiltonian $J(\xi)$ has to be linear in the electromagnetic momentum E and analogous to Example 2.137 we hence have to have

$$J(\xi) = \int_Q E \cdot \nabla\xi \, dq. \tag{2.313}$$

From the definition of the momentum map it has to hold

$$\langle \mathbf{J}(A, -E), \xi \rangle = -\int_Q E \cdot \nabla\xi \, dq \tag{2.314}$$

and integration by parts[178] for fields vanishing at the boundary or decaying sufficiently fast at infinity yields

$$\langle \mathbf{J}(A, -E), \xi \rangle = -\int_Q (\nabla \cdot E)\, \xi \, dq. \tag{2.315}$$

Comparing to Eq. 2.314, the momentum map can be read off directly as

$$\mathbf{J}(A, -E) = \nabla \cdot E \ \in \mathfrak{g}^* \cong \mathcal{F}(\mathbb{R}^3) \tag{2.316}$$

The divergence freeness of the electric field was the fourth of Maxwell's equations that we could not yet derive in Example 2.128, and it is interesting to observe that within a modern mathematical framework the constraint is closely related to the conservation of linear and angular momentum.

Similar to other results consider in Chapter 2.3.4.3, a simpler description of the momentum map is possible when the symplectic manifold (P, ω) is a cotangent bundle.

Proposition 2.102. *Let $\phi : \mathfrak{g} \times Q \to TQ : \xi \times q \to \xi_Q(q)$ be the action of a Lie algebra $\mathfrak{g}$ on a manifold Q, and by cotangent lifts on the cotangent bundle T^*Q with coordinates $(q, p) \in T^*Q$, and let $L : TQ \to \mathbb{R}$ be a $\mathfrak{g}$-invariant Lagrangian*

---

[178]We employ here $\int_{\mathcal{M}} \nabla U \cdot \vec{A}\, dV = \int_{\partial\mathcal{M}} (U\vec{A}) \cdot dA - \int_{\mathcal{M}} U \nabla \cdot \vec{A}\, dV$.

such that $\xi_Q[L] = 0$. Then the momentum Hamiltonian $J : \mathfrak{g} \to \mathcal{F}(T^*Q)$ satisfies

$$J = \langle \mathbb{F}L, \xi_Q(q) \rangle$$

and the momentum map $\mathbf{J} : T^*Q \to \mathfrak{g}^*$ is given by

$$\langle \mathbf{J}(q, p), \xi \rangle = \langle \mathbb{F}L, \xi_Q(q) \rangle = \langle p, \xi_Q(q) \rangle.$$

Remark 2.169. The above result is sometimes known as Noether's formula for cotangent bundles.[179]

Example 2.140. Let us consider again a particle under the action of $(\mathbb{R}^3, +)$ for which we already computed the momentum map in Example 2.137. The infinitesimal generator ξ_Q for the spatial part of the dynamics, corresponding to $\dot{q}$, is given by ξ_Q and by the definition of the action it does not depend on the position $q \in Q$. Hence, by Proposition 2.102 the momentum Hamiltonian is

$$J(\xi)(q, p) = p \cdot \xi_Q \tag{2.317}$$

and the momentum map can be read off directly,

$$\mathbf{J}(q, p) = p \ \in \mathfrak{g}^* \cong \mathbb{R}^3 \tag{2.318}$$

which agrees with our calculation in Example 2.137.

Example 2.141. Consider again a particle in $\mathbb{R}^3$ under the action of SO(3) as in Example 2.138. The infinitesimal generator $\xi_Q(q)$ for the spatial component is

$$\xi_Q(q) = \xi \, q = \check{\xi} \times q \tag{2.319}$$

where $\xi \in \mathfrak{so}(3)$ is a skew-symmetric matrix and the hat map of the representation $\check{\xi} \in \mathbb{R}^3$, cf. Example 2.93. Hence, using elementary identities for the cross product, the momentum Hamiltonian is given by

$$J(\omega)(q, p) = p \cdot (\xi \, q) = p \cdot (\check{\omega} \times q) = \check{\omega} \, (q \times p) \tag{2.320}$$

and the momentum map can hence again be read off as

$$\mathbf{J} = q \times p \tag{2.321}$$

which is the usual expression for the angular momentum and agrees with the previous calculation in Example 2.138.

The construction of the momentum map and Hamiltonian can be described by the diagram in Fig. 2.46. The mappings $X \mapsto X_H$ and $\xi \mapsto \xi_P$ are Lie algebra

---

[179]Holm, Schmah, and Stoica, *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*, p. 249.
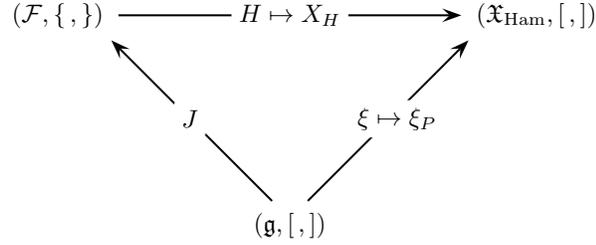
Figure 2.46: Lie algebra homomorphisms for the maps in the definition of the momentum map. The momentum Hamiltonian J provides a homomorphism from the Lie algebra $(g, [\,,])$ into the function algebra $(\mathcal{F}(P), \{\,,\})$ when the momentum map is equivariant.

(anti-)homomorphisms, cf. Proposition 2.96, but the momentum Hamiltonian $J : \mathfrak{g} \to \mathcal{F}(P)$ does in general not have this property, and one has

$$J([\xi, \eta]) - \{J(\xi), J(\eta)\} \equiv \Sigma(\xi, \eta) \tag{2.322}$$

for $\xi, \eta \in \mathfrak{g}$. $\Sigma(\xi, \eta)$ is known as 2-cocycle of $\mathfrak{g}$ and only when $\Sigma(\xi, \eta) = 0$ then J provides a Lie algebra homomorphism from the Lie algebra $(g, [\,,])$ into the function algebra $(\mathcal{F}(P), \{\,,\})$. This special but important case motivates the following definition which employs the notion of equivariance from Def. 2.195.

Definition 2.224. *Let $(G, \mathfrak{g})$ be a Lie group acting on the symplectic manifold (P, ω) canonically, and let the action $\varphi : G \times P \to P$ admit a momentum map $\mathbf{J}$. Then $\mathbf{J} : P \to \mathfrak{g}^*$ is **equivariant** when*

$$\mathrm{Ad}^*_{g^{-1}} \circ \mathbf{J} = \mathbf{J} \circ \varphi_g$$

and the following diagram commutes

$$
\begin{array}{ccc}
P & \xrightarrow{\ \mathbf{J}\ } & \mathfrak{g}^* \\
{\scriptstyle \varphi_g}\downarrow & & \downarrow{\scriptstyle \mathrm{Ad}^*_{g^{-1}}} \\
P & \xrightarrow[\ \mathbf{J}\ ]{} & \mathfrak{g}^*
\end{array}
$$

*A momentum map $\mathbf{J}$ is **infinitesimally equivariant** when for all $\xi, \eta \in \mathfrak{g}$*

$$J([\xi, \eta]) = \{J(\xi), J(\eta)\}$$

or equivalently

$$T_z\mathbf{J} \cdot \eta_P(z) = -\mathrm{ad}^*_\eta(\mathbf{J}(z)).$$

The above definition employed the coadjoint action $\mathrm{Ad}^* : G \times \mathfrak{g}^* \to \mathfrak{g}^*$ of a Lie group on its algebra and its infinitesimal generator $\mathrm{ad}^* : \mathfrak{g} \times \mathfrak{g}^* \to \mathfrak{g}$ which were introduced in Chapter 2.3.3.2, and $T_z\mathbf{J} \cdot \eta_P(z)$ is to be understood as the directional derivative in direction $\eta_P(z)$.

Proposition 2.103. *Let a Lie group* $(G, \mathfrak{g})$ *act canonically on a symplectic manifold* (P, ω)*, and let* $\mathbf{J}$ *be an equivariant momentum map for the action. Then* $\mathbf{J}$ *is also infinitesimally equivariant for the action of* $\mathfrak{g}$ *on* P*.*

The importance of the notion of equivariance lies in the fact that it enables to describe the dynamics of a system with symmetry on a reduced phase space as the commuting diagram in Def. 2.224 already suggests. However, before we discuss reduced dynamics in more detail, we will characterize when a momentum Hamiltonian J exists, which was stated somewhat vaguely in the definition of the momentum map in Def. 2.223.

Theorem 2.36. *A canonical Lie algebra action on a symplectic manifold* (P, ω) *is Hamiltonian admitting a moment Hamiltonian* J *if and only if there is a Lie algebra homomorphism* $f : \mathfrak{g} \to \mathcal{F}(P)$ *such that* $X_{f(\xi)} = \xi_P$ *for all* $\xi \in \mathfrak{g}$*. If such a Lie algebra homomorphism* f *exists then* $J = f$ *and it defines an infinitesimally equivariant momentum map* $\mathbf{J}$*, and any such momentum map defines a Lie algebra homomorphism.*

Reduction   The purpose and objective of reduction is to obtain a "simplified" description of a mechanical system by removing the symmetries—or redundancies—of a system. In the following we will see how the reduced phase space for a system can be constructed and how dynamics on this space can be described.

In the modern, Hamiltonian formulation of Noether's theorem in Theorem 2.35 we saw that the image $\mathbf{J}(z) = \mu \in \mathfrak{g}^*$ of the momentum map in the dual Lie algebra $\mathfrak{g}^*$ is conserved under the flow of a G-invariant Hamiltonian $H : P \to \mathbb{R}$ when the Lie group G acts canonically on the symplectic manifold (P, ω) with $z \in P$. A corollary of the result is that for such a system also the pre-image

$$\mathbf{J}^{-1}(\mu) \subset P \tag{2.323}$$

is an invariant of the motion, that is a system with conserved quantity $\mu \in g^*$ will over time always stay within $\mathbf{J}^{-1}(\mu)$, cf. Fig.2.17. The pre-image $\mathbf{J}^{-1}(\mu) \subset P$ can already by considered as a reduced phase space, since for an arbitrary value μ it describes the subset of phase space P which is relevant when $\mu \in \mathfrak{g}^*$ is the value of the conserved quantity at some initial time. That this subset is well defined is established in the following proposition, cf. Def. 2.80 for the relevant notions.

Proposition 2.104. *Let* $(G, \mathfrak{g})$ *be a Lie group acting canonically on the symplectic manifold* (P, ω)*,* $\mathbf{J} : P \to \mathfrak{g}^*$ *be an equivariant momentum map for the action, and* $H : P \to \mathbb{R}$ *be a* G*-invariant Hamiltonian. When* $\mathbf{J}(z) \in \mathfrak{g}^*$ *is a regular value of* $\mathbf{J}$ *then*

$$\mathbf{J}^{-1}(\mu) \subset P$$

is a submanifold of P, and when $\mathfrak{g}^*$ is finite dimensional then $\mathbf{J}^{-1}(\mu)$ has codimension $\dim(\mathfrak{g}^*)$*, that is*

$$\dim(\mathbf{J}^{-1}(\mu)) = \dim(P) - \dim(\mathfrak{g}^*).$$

Although we fixed the value of the conserved quantity $\mu \in \mathfrak{g}^*$, the pre-image $\mathbf{J}^{-1}(\mu)$ is still not free of symmetries. For example, the momentum map for the action of SO(3) on $\mathbb{R}^3$ in Example 2.138 was given by $\mathbf{J}(q,p) = q \times p$. However, the system still possesses an SO(2) symmetry with respect to rotations around the fixed axis $\mu = q \times p$. These remaining symmetries of $\mathbf{J}^{-1}(\mu)$ are described by the stabilizer G_μ, that is the subgroup of G leaving μ invariant under the coadjoint action $\mathrm{Ad}^* : G \times \mathfrak{g}^* \to \mathfrak{g}^*$ on the dual Lie algebra $\mathfrak{g}^*$.[180] We hence obtain the following definition.

Definition 2.225. *Let $(G, \mathfrak{g})$ be a Lie group acting canonically on the symplectic manifold (P, ω), $\mathbf{J} : P \to \mathfrak{g}^*$ be an equivariant momentum map for the action, and $H : P \to \mathbb{R}$ be a G-invariant Hamiltonian. The **reduced phase space** P_μ at the regular value μ of the conserved quantity $\mathbf{J}(z) \in \mathfrak{g}^*$ is the quotient*

$$P_\mu = \mathbf{J}^{-1}(\mu)/G_\mu$$

of the pre-image $\mathbf{J}^{-1}(\mu) \subset P$ by the stabilizer G_μ, and the associated quotient map from the pre-image $\mathbf{J}^{-1}(\mu)$ to the reduced phase space is

$$\pi_\mu = \mathbf{J}^{-1}(\mu) \to P_\mu.$$

It is a corollary of Proposition 2.85 that the reduced phase space P_μ is a smooth manifold when the action of G_μ on $\mathbf{J}^{-1}(\mu)$ is free and proper, and we will assume this in the following.

Remark 2.170. The necessary modifications when μ is not a regular value of the momentum map or when the quotient $\mathbf{J}^{-1}(\mu)/G_\mu$ is not a smooth manifold are topics studied in singular reduction theory, which are however beyond the scope of the present thesis.[181]

Example 2.142. We considered before the example of the action of SO(3) on $T^*\mathbb{R}^3$ by cotangent lifts from $\mathbb{R}^3$, where $T^*\mathbb{R}^3$ can be considered as the phase space of a classical particle. From Example 2.138 we know that the momentum map for the system is $\mathbf{J}(q,p) = q \times p \in \mathfrak{so}^*(3)$ and $\mathfrak{so}(3) \cong \mathbb{R}^3$ through the hat map. The pre-image $\mathbf{J}^{-1}(\mu)$ is hence the set of all phase space points whose angular momentum is $\mu = (\mu^1, \mu^2, \mu^3) \in \mathbb{R}^3$, and which all have the same axis

[180]It was one of the contributions of Smale ("Topology and Mechanics I"; "Topology and Mechanics II") to reduction theory to recognize that the stabilizer G_μ characterizes the remaining symmetries in $\mathbf{J}^{-1}(\mu)$.

[181]See for example (Ortega and Ratiu, *Momentum Maps and Hamiltonian Reduction*) and the overview in (Marsden and Ratiu, *Mechanical Systems: Symmetry and Reduction*).
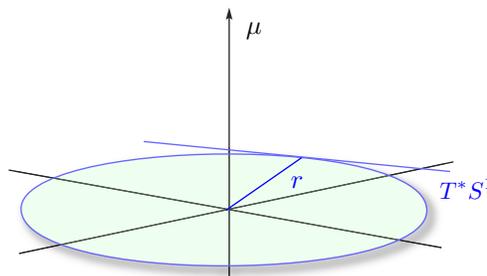
Figure 2.47: Pre-image $\mathbf{J}^{-1}(\mu)$ of the momentum map in phase space for a fixed angular momentum $\mu \in \mathfrak{so}^*(3)$ for the lifted action of SO(3) on $T^*\mathbb{R}^3$.

of rotation $\mu/\|\mu\|$. Since the axis of rotation is fixed, all such configurations lie in a plane with normal μ, and the pre-image of the momentum map is hence

$$\mathbf{J}^{-1}(\mu) = T^*S^1(r) \times \mathbb{R}^+ \tag{2.324}$$

where T^*S^1 represents the circular motion around μ and $\mathbb{R}^+$ corresponds to the radial direction, that is $r \in \mathbb{R}^+$, see Fig. 2.47. Also note that $\mathbf{J}^{-1}(\mu)$ is a subbundle of the cotangent bundle of the plane when considered in spherical coordinates. As required by the general theory in Proposition 2.104 we have

$$\dim\left(T^*S^1\right) \times \mathbb{R} = 2 + 1 = \dim\left(T^*\mathbb{R}^3\right) - \dim\left(\mathfrak{so}^*(3)\right) = 3. \tag{2.325}$$

The subgroup G_μ in SO(3) that leave $\mu \in \mathfrak{so}^*(3)$ invariant is given by all rotations around the axis μ, that is

$$G_\mu = \mathrm{SO}(2). \tag{2.326}$$

The reduced phase space for the lifted action of SO(3) on $\mathbb{R}^3$ is hence

$$P_\mu = \mathbf{J}^{-1}(\mu)/G_\mu = \left(T^*S^1(r) \times \mathbb{R}^+\right)/\mathrm{SO}(2) = \mathbb{R} \times \mathbb{R}^+ \tag{2.327}$$

since the quotient $S^1/\mathrm{SO}(2)$ is a single point. We can look at the reduced phase space also from a more intuitive perspective. When the particle is described in $\mathbb{R}^3$, then we can choose any of the equivalent orientations of the Cartesian coordinate systems, and when for convenience we chose the orientation that aligns the axis of rotation with the z axis, then we still have the freedom to choose any of the equivalent orientations of the radial coordinate system for the x-y plane in which the particle is then moving. The arbitrariness in the choices for the coordinate axes then correspond to the symmetries of the system.

Remark 2.171. We can think of the homogeneity of the Hamiltonian H on the cotangent bundle T^*Q as an $\mathbb{R}^+$ symmetry of the system, and the cosphere bundle $S^*Q = (T^*Q\backslash\{0\})/R^+$ is then the reduced phase space for the system, see also the discussion in Example 2.135.

Given $P_\mu = \mathbf{J}^{-1}(\mu)/G_\mu$, we would now like to describe the dynamics on the reduced phase space. Since we began with a symplectic manifold, the reduced dynamics are obtained from the symplectic reduction theorem.

Theorem 2.37 (Symplectic Reduction Theorem).[182] *Let $(G, \mathfrak{g})$ be a Lie group acting canonically on the symplectic manifold (P, ω), $\mathbf{J} : P \to \mathfrak{g}^*$ be an equivariant momentum map for the action, and $H : P \to \mathbb{R}$ be a G-invariant Hamiltonian. Then the **symplectic reduced space** (P_μ, ω_μ) is a symplectic manifold with the unique symplectic structure*

$$\pi_\mu^* \omega_\mu = i_\mu^* \omega,$$

where $i_\mu : \mathbf{J}^{-1}(\mu) \to P$ is the inclusion map.

The importance of the symplectic reduction theorem lies in the symplectic structure that it establishes on the reduced phase space P_μ, and that the symplectic structures on P and P_μ are consistent and related by the quotient map $\pi_\mu = \mathbf{J}^{-1}(\mu) \to P_\mu = \mathbf{J}^{-1}(\mu)/G_\mu$, which enables to employ the map π_μ to "drop" the dynamics from P to P_μ. The reduced dynamics are characterized in the following results.

Corollary 2.22. *Let the same assumptions as in Theorem 2.37 hold, and let $F, G \in \mathcal{F}(P_\mu)$ with $\hat{F}, \hat{G} \in \mathcal{F}(\mathbf{J}^{-1}(\mu))$ being defined by $\hat{F} = \pi_\mu^{-1}(F)$ and $\hat{G} = \pi_\mu^{-1}(G)$. Then*

$$\{\hat{F}, \hat{G}\} = \{F, G\}_\mu \circ \pi_\mu$$

where $\{\,,\,\}_\mu$ is canonical Poisson bracket induced by ω_μ.

The above corollary shows that also the Poisson structures on phase space P and its reduced counterpart P_μ at $\mu \in \mathfrak{g}^*$ are π_μ-related.

Definition 2.226. *Let the same assumptions as in Theorem 2.37 hold. Then the **reduced Hamiltonian** H_μ is*

$$H_\mu \circ \pi_\mu = H \circ i_\mu$$

*and it defines a **reduced Hamiltonian vector field** X_H^μ on P_μ with **reduced Hamiltonian flow** $\varphi_t^\mu : \mathbb{R} \times P_\mu \to P_\mu$.*

The reduced Hamiltonian dynamics on P_μ are related to the unreduced ones on P as follows.

Proposition 2.105. *Let the same assumptions as in Theorem 2.37 hold. The flow φ_t^μ on P_μ generated by the reduced Hamiltonian H_μ then satisfies*

$$\varphi_t^\mu \circ \pi_\mu = \pi_\mu \circ \varphi_t,$$

---

[182]The theorem is also known as Marsden-Weinstein reduction theorem, although it was established simultaneously by Marsden and Weinstein ("Reduction of Symplectic Manifolds with Symmetry") and Meyer ("Symmetries and Integrals in Mechanics").

where φ_t is the flow of the unreduced Hamiltonian system on $\mathbf{J}^{-1}(\mu) \subset P$, and the following diagram commutes

$$
\begin{array}{ccc}
\mathbf{J}^{-1} & \xrightarrow{\varphi_t} & \mathbf{J}^{-1} \\
\pi_\mu \downarrow & & \downarrow \pi_\mu \\
P_\mu & \xrightarrow[\varphi_t^\mu]{} & P_\mu
\end{array}
$$

Infinitesimally, also the Hamiltonian vector fields X_H^μ and X_H are π_μ-related satisfying

$$(T\pi_\mu)X_H = X_H^\mu \circ \pi_\mu.$$

The symplectic reduction theorem relates the symplectic structure on P to a reduced symplectic structure on the reduced phase space P_μ. When we start from a Poisson manifold, which might have a non-canonical Poisson structure that does not arise from a symplectic 2-form, then the following reduction theorem applies, which will also be useful in the sequel for Lie-Poisson reduction. The reader should consult Remark 2.156 for the necessary notions of Poisson geometry.

Theorem 2.38 (Poisson Reduction Theorem)**.** *Let G be a Lie group acting via Poisson maps on the Poisson manifold $(P, \{\,,\})$, and assume that the action is free and proper such that the quotient P/G is a smooth manifold with quotient projection $\pi : P \to P/G$. Then $(P/G, \{\,\}_r)$ is a Poisson manifold and π is a Poisson map defining the reduced Poisson structure by*

$$\{F, G\}_r \circ \pi = \{F \circ \pi, G \circ \pi\}$$

for functions $F, G \in \mathcal{F}(P/G)$.

The structure of the Poisson reduction theorem should be compared to the symplectic reduction theorem in Theorem 2.37.

Example 2.143. We discussed electromagnetic theory in the foregoing in Examples 2.125, 2.128, and 2.139, and we saw there that the theory has a gauge symmetry. Carrying out Poisson reduction for the bracket in Eq. 2.271 yields

$$\{F, G\} = \int_Q \left(\frac{\delta F}{\delta E} \cdot \left(\nabla \times \frac{\delta G}{\delta B} \right) - \frac{\delta G}{\delta E} \cdot \left(\nabla \times \frac{\delta F}{\delta B} \right) \right) dq \qquad (2.328)$$

and this bracket is known as Maxwell-Poisson bracket or Pauli-Born-Infeld bracket, since it was (re)-discovered by Pauli, Born, and Infeld.

Next to symplectic and Poisson reduction, various other types and variants of reduction for Hamiltonian systems such as cotangent bundle reduction, Lie-Poisson reduction, semi-direct product reduction, reduction by stages, reduction without momentum maps, multi-symplectic reduction, discrete reduction,

reduction for systems with non-holonomic constraints and Dirac constraints, and singular reduction exist, and today also Lagrangian reduction is a well developed subject.[183] In the next section Lie-Poisson reduction will be studied in more detail.

2.3.5.2 Reduction for Lie-Poisson Systems

In this section we will study Lie-Poisson reduction in more detail, that is reduction for systems whose configuration space is a Lie group, and since a group is acting on a manifold that has a group structure itself, there is a particularly rich mathematical structure. In fact, it is the abundance of structure that makes it at times challenging to work with these systems.

Definition 2.227. *A **Lie-Poisson system** is a mechanical system whose configuration space is a Lie group* $(G, \mathfrak{g})$.

As mentioned before, classical examples of Lie-Poisson systems are the Euler top and the ideal Euler fluid, cf. Table 2.1. Systems where the group structure has been discovered more recently are the Korteweg-de Vries equation and the Maxwell-Vlasov and Vlasov-Poisson systems. In the following, we will develop Lie-Poisson theory for an arbitrary, possibly infinite dimensional Lie group $(G, \mathfrak{g})$.

Group Actions for Lie-Poisson Systems For Lie-Poisson systems, the group action is defined by the left and right translation maps

$$L : G \times G \to G \; : \; L_h\, g = h \circ g \tag{2.329a}$$

$$R : G \times G \to G \; : \; R_h\, g = g \circ h \tag{2.329b}$$

that is the group is acting by an automorphism on itself. The tangent map for the action is

$$TL : G \times TG \to TG : (TL_h)(g, X_g) = (h\, g, \; (T_g L_h)\, X_g) \tag{2.330a}$$

$$TR : G \times TG \to TG : (TR_h)(g, X_g) = (g\, h, (T_g R_h)\, X_g) \tag{2.330b}$$

and it should be recalled from Chapter 2.3.3 that these define left and right invariant vector fields on G, respectively, that is the so defined vector fields satisfy $L_h^* X_g = (T_g L_h) X_g = X(h \circ g)$ and $R_h^* X_g = (T_g R_h) X_g = X(g \circ h)$, cf. Def. 2.177. By the definition of the cotangent lift, cf. Def. 2.95, one obtains for the action of G on its cotangent bundle T^*G that

$$T^*L : G \times T^*G \to T^*G : (T^*L_{h^{-1}})(g, \alpha_g) = (L_h\, g, (T_{L_h g}^*\, L_{h^{-1}})\, \alpha_g) \tag{2.331a}$$

$$T^*R : G \times T^*G \to T^*G : (T^*R_{h^{-1}})(g, \alpha_g) = (R_h\, g, (T_{R_h g}^*\, R_{h^{-1}})\, \alpha_g) \tag{2.331b}$$

---

[183]See (Ortega and Ratiu, *Momentum Maps and Hamiltonian Reduction*; Marsden and Ratiu, *Mechanical Systems: Symmetry and Reduction*) for overviews of different kinds of reduction.

and the conventions are chosen such that the cotangent lift of a left action yields again a left action, and analogous for right actions. By construction, the cotangent lifted action is canonical, preserving the natural symplectic 2-form on T^*G, cf. Example 2.119.

The systems of interest for us have a left or right invariant Hamiltonian, and hence the conserved quantities are obtained using the momentum maps for the translation actions.[184] Interestingly, for a left invariant Hamiltonian the momentum map associated with the right translation action has to be employed, and analogously for a right invariant Hamiltonian, and we will see the reason shortly after we computed the momentum maps for Lie-Poisson actions, which follows next.

Lie-Poisson Momentum Maps As we discussed in the previous section, a vital ingredient for obtaining a reduced system is the momentum map. In Proposition 2.102 we established that for actions lifted from a configuration manifold Q to the associated cotangent bundle T^*Q the momentum map is given by

$$\langle \mathbf{J}(q,p), \xi \rangle = \langle p, \xi_Q(q) \rangle \tag{2.332}$$

where ξ_Q denotes the part of the infinitesimal generator on the base space Q. We could employ the proposition directly to obtain the momentum map for the action of G on T^*G. However, we will only use that $\mathbf{J} : T^*G \to \mathfrak{g}^*$ solely depends on ξ_G, the infinitesimal generator on the base G of T^*G, and compute the momentum map "by hand", which will provide some insight into the resulting expression. Similar to the examples in the previous sectin, ξ_G is obtained using the definition of the infinitesimal generator with the exponential map. For the left action $L_h\, g = h \circ g$ and the acting group element h of the form $h(t) = \exp(t\xi)$ for ξ, this yields

$$\xi_G^L = \frac{d}{dt} h_\xi(t) \circ g \bigg|_{t=0} = \frac{d}{dt} \exp(t\,\xi) \circ g \bigg|_{t=0} = \xi\, g = (T_e R_g)\, \xi. \tag{2.333}$$

where $\xi\, g$ should be understood as the Lie algebra action of $\xi \in \mathfrak{g}$ on $g \in G$. Hence, the infinitesimal generator ξ_G^L for the left action is the right invariant vector field $(T_e R_g)\, \xi$ generated by the Lie algebra element $\xi \in \mathfrak{g}$. Analogously, for the right action $R_h\, g = g \circ h$ one obtains

$$\xi_G^R = g \circ \frac{d}{dt} h_\xi(t) \bigg|_{t=0} = g \circ \frac{d}{dt} \exp(t\,\xi) \bigg|_{t=0} = g\, \xi = (T_e L_g)\, \xi \tag{2.334}$$

and the infinitesimal generator ξ_G^R for the right action is given by the left invariant vector field $(T_e L_g)\, \xi$ generated by $\xi \in \mathfrak{g}$.

---

[184]Certain systems such such as gravitating fluids are both left *and* right invariant (Marsden et al., "Hamiltonian Systems with Symmetry, Coadjoint Orbits and Plasma Physics", p. 11) but we will disregard this special case from our considerations.

Since T^*G is a cotangent bundle, the Hamiltonian vector field is given by Hamilton's equations and hence the infinitesimal generator ξ_G has to coincide with the spatial velocity $\dot{q}$ defined by them, that is for the yet unknown momentum Hamiltonian J we have to have

$$\xi_G = \frac{\partial J}{\partial \alpha_g} \tag{2.335}$$

where $\alpha_g \in T_g^*G$ is the momentum covector on the cotangent bundle of the group. Comparing the expressions for ξ_G^L and ξ_G^R with Eq. 2.335 shows that the momentum Hamiltonian $J(\xi : \mathfrak{g} \to \mathcal{F}(T^*G)$ has to be linear in the momentum so that

$$J_L(\xi) = \xi_G^L \cdot \alpha_g = ((T_e R_g)\,\xi) \cdot \alpha_g \tag{2.336a}$$

$$J_R(\xi) = \xi_G^R \cdot \alpha_g = ((T_e L_g)\,\xi) \cdot \alpha_g \tag{2.336b}$$

But by the definition of the pullback it holds

$$((T\varphi)X) \cdot \alpha_g = (\varphi^*\alpha_g)(X), \tag{2.337}$$

that is the tangent map can be formulated as a pullback, so that the momentum Hamiltonians $J_L(\xi)$ and $J_R(\xi)$ can be written as

$$J_L(\xi) = ((T_e R_g)\,\xi) \cdot \alpha_g = ((T_e^* R_g)\,\alpha_g) \cdot \xi \tag{2.338a}$$

$$J_R(\xi) = ((T_e L_g)\,\xi) \cdot \alpha_g = ((T_e^* L_g)\,\alpha_g) \cdot \xi. \tag{2.338b}$$

Comparing the above equations to the definition of the momentum map, $\langle \mathbf{J}(z), \xi \rangle = J(\xi)(z)$, we obtain the left and right momentum maps for Lie-Poisson systems, and these are provided in the following proposition.

Proposition 2.106. *The momentum map for a Lie-Poisson system when the Lie group G is acting by the left translation map is*

$$\mathbf{J}_L(g, \alpha_g) = (T_e^* R_g)\,\alpha_g \ : \ T^*G \to \mathfrak{g}_+^*$$

and when G is acting by the right translation map it is

$$\mathbf{J}_R(g, \alpha_g) = (T_e^* L_g)\,\alpha_g \ : \ T^*G \to \mathfrak{g}_-^*.$$

Since for a Lie-Poisson system there are two ways to pass to the dual Lie algebra we wrote $\mathfrak{g}_+^*$ and $\mathfrak{g}_-^*$ in the above definition, and the difference will be of considerable importance in the following. It should again be noted that the momentum map $\mathbf{J}_L$ for the left translation action is given by the cotangent lift of the right translation map, and analogously $\mathbf{J}_L$ is given by the cotangent lift of the left translation map. This implies that $\mathbf{J}_L(g, \alpha_g)$ is left invariant while $\mathbf{J}_R(g, \alpha_g)$ is right invariant, that is

$$R_h \circ \mathbf{J}_L(g, \alpha_g) = \mathbf{J}_L(R_h \circ (g, \alpha_g)) = \mathbf{J}_L(g, \alpha_g) \tag{2.340a}$$

$$L_h \circ \mathbf{J}_R(g, \alpha_g) = \mathbf{J}_R(L_h \circ (g, \alpha_g)) = \mathbf{J}_R(g, \alpha_g) \tag{2.340b}$$
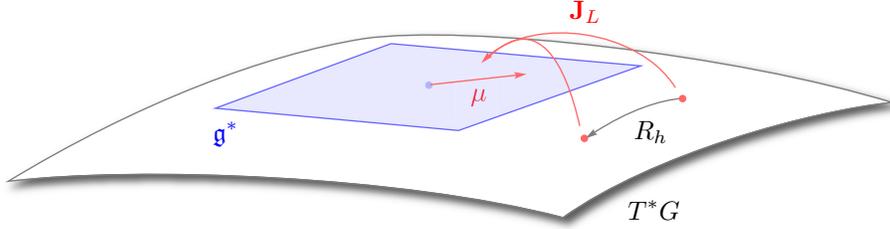
Figure 2.48: The left momentum map $\mathbf{J}_L : T^*G \to \mathfrak{g}_+^*$ is invariant under the right action R_h of G on its cotangent bundle T^*G by cotangent lifts.

and where the left and right actions on (g, α_g) are given by the cotangent lift of the translation action in Eq. 2.331, see Fig. 2.48. Indeed, the left action on the cotangent bundle by an element $h \in G$ is

$$L_h \circ (g, \alpha_g) = (L_h\, g, (T^*_{L_h g}\, L_{h^{-1}})\, \alpha_g) = (hg, (T^*_{hg}\, L_{h^{-1}})\, \alpha_g) \qquad (2.341a)$$

while the tangent map $\mathbf{J}_R$ of the right action at hg is

$$\mathbf{J}_R = T^*_e L_{hg} = T^*_e (L_h \circ L_g) = (T^*_e L_g) \circ (T^*_g L_h) \qquad (2.341b)$$

since the pullback "inverts the direction", cf. Fig. 2.33. Combining Eq. 2.341a and Eq. 2.341b we hence obtain

$$L_h \circ \mathbf{J}_R(g, \alpha_g) = (T^*_e L_g) \circ (T^*_g L_h)((T^*_{hg}\, L_{h^{-1}})\, \alpha_g) \qquad (2.341c)$$

and since $(T^*_g L_h)$ and $(T^*_{hg}\, L_{h^{-1}})$ are inverse maps we have

$$L_h \circ \mathbf{J}_R(g, \alpha_g) = (T^*_e L_g)\, \alpha_g \qquad (2.341d)$$

which shows our claim since the right hand side is just the expression or the right Lie-Poisson momentum map in Proposition 2.106. An analogous derivation shows the right invariance of $\mathbf{J}_L$.

With the Lie-Poisson momentum maps in Proposition 2.106, it follows from Noether's theorem that when a Hamiltonian is invariant under the right action of a group then the conserved quantities are obtained using $\mathbf{J}_R$ in the left dual Lie algebra $\mathfrak{g}_-^*$, while when a Hamiltonian is invariant under the left action then the conserved quantities are obtained using $\mathbf{J}_L$ in the right dual Lie algebra $\mathfrak{g}_+^*$. For example, for the Euler top the Hamiltonian is left invariant so that the image of $\mathbf{J}_L$ in $\mathfrak{so}(3)$ represents the conserved quantities, which, as we see shortly, corresponds to the spatial angular momentum whose conservation is well known. The following proposition is an immediate consequence of the fact that the Lie-Poisson momentum maps are obtained from a cotangent lift.

Proposition 2.107. *The Lie-Poisson momentum maps $\mathbf{J}_L$ and $\mathbf{J}_R$ are equivariant.*

The proposition also implies that the momentum maps are Poisson maps. For simplicity, in the following we will only consider a left Lie-Poisson system, that is when the Lie group G is acting by the left translation map on itself, and we will point out the differences when appropriate.

Lie-Poisson Reduced Phase Space With the momentum map, we can determine the reduced phase space $P_\mu = \mathbf{J}^{-1}(\mu)/G_\mu$. Since the dual Lie algebra $\mathfrak{g}^*$ is isomorphic to the cotangent space $T_e^* G$ of G at the identity, that is $\mathfrak{g}^* \cong T_e^* G$, the pre-image $\mathbf{J}_L^{-1}(\mu) \subset T^* G$ of the momentum map for fixed $\mu \in \mathfrak{g}^*$ on phase space $T^* G$ is the space of right invariant 1-forms $\alpha_\mu \in T^* G$ defined by

$$(T_e^* R_g)\, \alpha_\mu(g) = \mu \tag{2.342}$$

which is equivalent to

$$\mathbf{J}_L(\alpha_\mu(g)) = \mu, \tag{2.343}$$

cf. Proposition 2.106. By construction, there is exactly one 1-form in each fiber of $\mathbf{J}_L^{-1}(\mu)$, and since $\mathfrak{g}^* \cong T_e^* G$ and all tangent spaces $T_g G$ are identical up to translation, we have for fixed $\mu \in \mathfrak{g}^*$ that $G \cong \mathbf{J}_L^{-1}(\mu) \subset T^* G$. Hence, the pre-image $\mathbf{J}_L^{-1}(\mu)$ can be identified with the group G itself and we choose the isomorphism to be $(g, \alpha_\mu) \mapsto g^{-1}$, which will be convenient later.

With the pre-image $\mathbf{J}_L^{-1}(\mu)$, we have to determine the quotient space $\mathbf{J}_L^{-1}(\mu)/G_\mu = G/G_\mu$. By definition, the stabilizer G_μ of $\mu \in \mathfrak{g}^*$ is

$$G_\mu = \left\{ g \in G \mid \mu = \mathrm{Ad}_g^*(\mu) \right\} \tag{2.344}$$

where Ad^* is again the coadjoint action of the Lie group on its dual Lie algebra $\mathfrak{g}^*$. Taking the quotient G/G_μ is identical to identifying all elements in the stabilizer with the identity e of G. We are then left in G/G_μ with the elements that act effectively on μ, that is elements $g \in G$ such that

$$\mu \neq \nu = \mathrm{Ad}_g(\mu) \in \mathfrak{g}^* \tag{2.345}$$

for $g \in G$. Turning the point of view around, and keeping in mind that Ad^* is a representation in the sense of Def. 2.185, for the quotient G/G_μ we are looking for $\nu \in \mathfrak{g}^*$ such that there exists $g \in G$ with $\nu = \mathrm{Ad}_g(\mu)$. But by definition, these dual Lie algebra elements $\nu \in \mathfrak{g}^*$ form the coadjoint orbit

$$\mathcal{O}(\mu) = \left\{ \nu \in \mathfrak{g}^* \mid \nu = \mathrm{Ad}_{g^{-1}}^*(\mu)\,,\ g \in G \right\}, \tag{2.346}$$

cf. Def. 2.187. For the reduced phase space for the action of G on its cotangent bundle $T^* G$ we therefore have

$$P_\mu = \mathbf{J}_L^{-1}(\mu)/G_\mu \cong G/G_\mu \cong \mathcal{O}(\mu) \tag{2.347}$$

with quotient projection

$$\pi_\mu(\alpha_\mu(g)) \mapsto \mathrm{Ad}_g^*(\mu) \tag{2.348}$$

where $\alpha_\mu(g)$ is as before the right invariant 1-form at g defined by $\mu \in \mathfrak{g}^*$, and we used that $(g^{-1})^{-1}$ which motivated our previous choice in the identification $\mathbf{J}_L^{-1}(\mu) \cong G$. From the above argument, it is evident that an analogous construction holds for $\mathbf{J}_R$ with the only difference being that the pre-image $\mathbf{J}^{-1}(\mu)$ is then given by left invariant 1-forms.

Lie-Poisson Reduced Dynamics Given the reduced phase space $P_\mu \cong \mathcal{O}(\mu)$ in Eq. 2.347 we can employ the symplectic reduction theorem in Theorem 2.37. We then immediately know that the coadjoint orbits are symplectic manifold, and it can be shown that the reduced symplectic 2-form ω_μ on the them—known as Kirillov-Kostant-Souriau structure—is given by[185]

$$\omega_\mu^\pm \left(\mathrm{ad}_\xi^*(\mu), \mathrm{ad}_\eta^*(\mu) \right) = \pm \langle \mu, [\xi, \eta] \rangle \in \Omega^2(\mathfrak{g}_\pm^*) \tag{2.349}$$

where $\mu \in \mathfrak{g}^*$ and the sign is determined by how one passes to the Lie algebra. The infinitesimal coadjoint actions $\mathrm{ad}_\xi^*(\mu)$ and $\mathrm{ad}_\xi^*(\eta)$ are tangent vectors in $T_\mu \mathfrak{g}^*$ for the flow generated by the Lie algebra elements $\xi, \nu \in \mathfrak{g}$, which can be seen by the argument in the following remark.

Remark 2.172. Let $g(t) : [a, b] \to G$ be an arbitrary curve on the Lie group $(G, \mathfrak{g})$ of the form $g(t) = \exp(t\xi)$. Then for any $\mu \in \mathfrak{g}^*$, the curve $g(t)$ induces a curve $\mu(t) = \mathrm{Ad}_{g(t)^{-1}}(\mu)$ on the dual Lie algebra $\mathfrak{g}^*$ by the coadjoint action. For an arbitrary $\eta \in \mathfrak{g}$ we whence have by the definition of the Ad^* action that

$$\langle \mu(t), \eta \rangle = \langle \mathrm{Ad}_{g(t)^{-1}}^*(\mu), \eta \rangle \tag{2.350a}$$

$$= \langle \mu, \mathrm{Ad}_{g(t)^{-1}}(\eta) \rangle. \tag{2.350b}$$

Differentiating both sides with respect to time at $t = 0$ yields

$$\langle \mu'(t), \eta \rangle = \langle \mu, \mathrm{ad}_\xi(\eta) \rangle. \tag{2.350c}$$

Using the definition of the ad^* operator we obtain

$$\langle \mu'(t), \eta \rangle = \langle \mathrm{ad}_\xi^*(\mu), \eta \rangle \tag{2.350d}$$

which shows that $\mu'(t) = \mathrm{ad}_\xi^*(\mu)$, and the tangent space to the coadjoint orbit $\mathcal{O}_\mu$ containing μ is hence given by

$$T_\mu \mathcal{O}_\mu = \left\{ \mathrm{ad}_\xi^* \mu \mid \xi \in \mathfrak{g} \right\} \subset T\mathfrak{g}^*. \tag{2.351}$$

By Def. 2.212, the reduced symplectic 2-form $\omega_\mu^\pm$ in Eq. 2.349 defines a canonical Poisson bracket on the coadjoint orbits $\mathcal{O}_\mu$ by

$$\{F, G\}_\mu = \omega_\mu^\pm(X_F, X_G) = \omega_\mu^\pm \left(\mathrm{ad}_\xi^*(\mu), \mathrm{ad}_\eta^*(\mu) \right) \tag{2.352}$$

---

[185]The existence of a natural symplectic structure on a dual Lie algebra was established independently in the 1960s by Kirillov, Kostant, and Souriau, see (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, p. 373) for historical remarks, and that it is a corollary of the symplectic reduction theorem was shown by Marsden and Weinstein ("Reduction of Symplectic Manifolds with Symmetry") which also provided an explanation for the existence of the structure.

To find a concrete expression for the bracket, we have to determine the Hamiltonian functions corresponding to the Hamiltonian vector fields $\mathrm{ad}^*_\xi(\mu)$ and $\mathrm{ad}^*_\eta(\mu)$, and we have to write them using the right hand side of Eq. 2.349 which provides the computational form of $\omega^\pm_\mu$. The definition of the reduced symplectic form $\omega^\pm_\mu$ can be written as

$$\omega^\pm_\mu \left(\mathrm{ad}^*_\xi(\mu), \mathrm{ad}^*_\eta(\mu) \right) = \pm \langle \mu, [\xi, \eta] \rangle \tag{2.353a}$$

$$\mathrm{i}_{\mathrm{ad}^*_\xi(\mu)} \, \omega^\pm_\mu \left(\mathrm{ad}^*_\eta(\mu) \right) = \mp \langle \mathrm{ad}_\eta(\xi), \mu \rangle \tag{2.353b}$$

$$\langle \mathrm{i}_{\mathrm{ad}^*_\xi(\mu)} \omega^\pm_\mu, \mathrm{ad}^*_\eta(\mu) \rangle = \mp \langle \xi, \mathrm{ad}^*_\eta(\mu) \rangle \tag{2.353c}$$

where the last equation is the pairing $\langle, \rangle$ of a 1-form and an arbitrary vector $\mathrm{ad}^*_\eta(\mu)$, that is Eq. 2.353c can also be written as

$$\mathrm{i}_{\mathrm{ad}^*_\xi(\mu)} \omega^\pm_\mu = \mp \xi \tag{2.353d}$$

and by the definition of the Hamiltonian vector field

$$\mathrm{i}_{X_H} \omega = \mathrm{d} H \tag{2.354}$$

we see that ξ has to correspond to $\mathrm{d} H$ with associated Hamiltonian vector field $\mathrm{ad}^*_\xi(\mu)$, where μ just denotes the location where the vector field is evaluated. The Lie algebra $\mathfrak{g}$ is possibly infinite dimensional. Hence, the functional derivative $\delta F/\delta\mu$ defined by

$$\mathrm{D} F \cdot \nu = \left\langle \frac{\delta F}{\delta \mu}, \nu \right\rangle, \tag{2.355}$$

where $F \in \mathcal{F}(\mathfrak{g}^*)$ and $\nu \in \mathfrak{g}^*$, has to be employed for the differential which is needed in the definition of the Hamiltonian vector field. Since the functional derivative $\delta F/\delta\mu$ is an element in the Lie algebra $\mathfrak{g}$,[186] we see from Eq. 2.353d that

$$\xi = \frac{\delta F}{\delta \mu} \approx \mathrm{d} F. \tag{2.356}$$

Hence, the canonical Poisson bracket on the coadjoint orbits is

$$\{F, G\}_\mu = \omega^\pm_\mu \left(\mathrm{ad}^*_{\frac{\delta F}{\delta \mu}}(\mu), \mathrm{ad}^*_{\frac{\delta G}{\delta \mu}}(\mu) \right) = \pm \left\langle \mu, \left[\frac{\delta F}{\delta \mu}, \frac{\delta G}{\delta \mu} \right] \right\rangle \tag{2.357}$$

The above bracket is only well defined on each coadjoint orbit $\mathcal{O}_\mu \subset \mathfrak{g}^*$. However, $\mathfrak{g}^*$ foliates into symplectic leaves and on each leaf there is a non-degenerate bracket, so that $\{,\}_\mu$ can in fact be extended to all $\mathfrak{g}^*$ as the disjoint union of the bracket on each orbit. We hence have established the following theorem.

---

[186] We identified here the double dual $\mathfrak{g}^{**}$ with $\mathfrak{g}$, that is we assumed that $\mathfrak{g}^*$ is reflexive. In the infinite dimensional setting this might not necessarily be satisfied.

Theorem 2.39. *Let $\mathfrak{g}$ be a Lie algebra with dual Lie algebra $\mathfrak{g}^*$, and let $\langle\,,\rangle$ be the natural pairing between $\mathfrak{g}$ and $\mathfrak{g}^*$. Then for $F, G \in \mathcal{F}(\mathfrak{g}^*)$ and $\mu \in \mathfrak{g}^*$ there exists a natural Poisson bracket on $\mathfrak{g}^*$ given by*

$$\{F, G\}(\mu) = -\left\langle \mu, \left[\frac{\delta F}{\delta \mu}, \frac{\delta G}{\delta \mu}\right]\right\rangle$$

and the bracket is known as **Lie-Poisson bracket**.[187]

Remark 2.173. Next to the construction from a reduced phase space, the Lie-Poisson bracket can also be obtained by more direct means and without reference to the symplectic structure on $\mathfrak{g}^*$.[188]

The importance of the Lie-Poisson bracket lies in its connection to the reduced phase space for Lie-Poisson systems and that it is naturally defines.

Example 2.144. For the Euler top with its SO(3) symmetry, the Lie-Poisson bracket is given by n

$$F, G(\Pi) = -\Pi \cdot (\nabla F \times \nabla G) \tag{2.358}$$

where $F, G \in \mathcal{F}(\mathfrak{so}^*_-(3))$, and $\Pi \in \mathfrak{so}^*_-(3)$ is the body angular momentum of the top.

Example 2.145. For an ideal Euler fluid the symmetry group is $\mathrm{Diff}_\mu(Q)$, the group of volume preserving diffeomorphisms, see Table 2.1 and the discussion in the next section. The Lie-Poisson bracket for functions $F(\mathfrak{g}^*_+)$ for the system is

$$\{F, G\} = \int_Q \left\langle \omega, \left[\frac{\delta F}{\delta \omega}, \frac{\delta G}{\delta \omega}\right]\right\rangle dq \tag{2.359}$$

where $F, G \in \mathcal{F}(\mathfrak{g}^*)$, $\omega \in \mathfrak{g}^*_+$ is the vorticity of the fluid.

The Lie-Poisson bracket together with the expression for the Hamiltonian vector field on $\mathfrak{g}^*$ that we derived previously yields the following theorem, cf. Def. 2.226 and Proposition 2.105.

Theorem 2.40. *Let $H : T^*G \to \mathbb{R}$ be the left invariant Hamiltonian for a Lie-Poisson system, and $H_- \in \mathcal{F}(\mathfrak{g}^*)$ be the induced Hamiltonian on $\mathfrak{g}^*_-$. Then the reduced Hamiltonian vector field $X_H^- \in T\mathcal{O}_\mu$ yields dynamics on the coadjoint orbits $\mathcal{O}_\mu$ given by*

$$\dot{\mu} = X_H^-(\mu) = \mathrm{ad}^*_{\frac{\delta H_-}{\delta \mu}}(\mu)$$

---

[187]The Lie-Poisson bracket appeared already in Lie's work (Lie, *Theorie der Transformationsgruppen*) but its significance was not realized until 80 years later in the work by Kirillov, Kostant, Souriau, Marsden, and Weinstein.

[188]For various direct proof of the Lie-Poisson reduction theorem see (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, Chapter 13) and for an explicit proof for infinite dimensional Lie groups see (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems – Online Supplement*).

and the Lie-Poisson evolution equations are

$$\dot{F}(\mu) = \{F, H_-\}(\mu)$$

for $F \in \mathcal{F}(\mathfrak{g}^*_-)$, where $\{\,,\,\}_-$ is the negative Lie-Poisson bracket. The coadjoint orbit $\mathcal{O}_\mu \subset \mathfrak{g}^*$ of $\mu \in g^*_-$ is invariant under these dynamics.

Remark 2.174. For right invariant systems the minus signs in the above theorem have to be replaced by plus signs.

Example 2.146.[189] Let $G = \mathrm{SO}(3)$ act on $\mathbb{R}^3$ with the usual rotation action. The configuration space of the system can then be identified with the group $\mathrm{SO}(3)$, since every state of the system can be described by a rotation with respect to some reference configuration, and the system is Lie-Poisson. Identifying $\mathfrak{so}(3)$ with $\mathbb{R}^3$ using the hat map, the dual Lie algebra $\mathfrak{so}^*(3)$ can also be identified with $\mathbb{R}^3$ using the usual dot product in Euclidean space, although one should be careful to distinguish the space the group is acting on and the Lie algebra $\mathfrak{so}(3)$ and its dual $\mathfrak{so}^*(3)$ since all three spaces are isomorphic to $\mathbb{R}^3$. The coadjoint orbits $\mathcal{O}(\mu)$ are then spheres $S^2(r)$ centered at the origin, and including the origin, and the symplectic structure on each orbit is given by the usual area element $d\omega$ which, by definition, is non-degenerate, see also Example 2.117. Since the area element $d\omega$ is invariant under rotation, the coadjoint action is canonical, and this also holds for the group action which also preserves the area element.

Theorem 2.40 on the reduced Hamiltonian dynamics for Lie-Poisson systems provides a description of the Hamiltonian dynamics on the orbits, and the Hamiltonian vector field is given by the infinitesimal generator of the coadjoint action

$$X_H^\mu = \mathrm{ad}^*_{\frac{\delta H_\mu}{\delta \mu}} \mu. \tag{2.360}$$

where $\mu \in \mathfrak{g}^*$ and in our case $\mathfrak{g}^* = \mathfrak{so}^*(3)$. For $\mathrm{SO}(3)$, the generator of the coadjoint action is given by[190]

$$\mathrm{ad}^*_\xi \omega = \xi \times \omega \tag{2.361}$$

where $\xi \in \mathfrak{so}(3) \cong \mathbb{R}^3$ and $\omega \in \mathfrak{so}^*(3) \cong \mathbb{R}^3$. We also know that the Hamiltonian vector field X_H has to coincide with the infinitesimal generator $\xi_{\mathbb{R}^3}$ of the $\mathrm{SO}(3)$ action,

$$\xi_{\mathbb{R}^3} = \xi \times \omega, \tag{2.362}$$

cf. Example 2.138. Comparing the above equations we see that the reduced Hamiltonian H_μ has to satisfy

$$\frac{\delta H_\omega}{\delta \omega} = \xi. \tag{2.363}$$

---

[189]The example was inspired by (Arnold and Khesin, *Topological Methods in Hydrodynamics*, Example 6.3) and (Polterovich, *The Geometry of the Group of Symplectic Diffeomorphisms*, Example 1.4.H).

[190]Holm, *Geometric mechanics: Rotating, translating and rolling*, Chapter 5.1.3.

$$\mathrm{Orb}(\mu)$$

$$\mathfrak{g}^*$$

$$\pi_\mu$$

$$e$$

$$T^*G$$
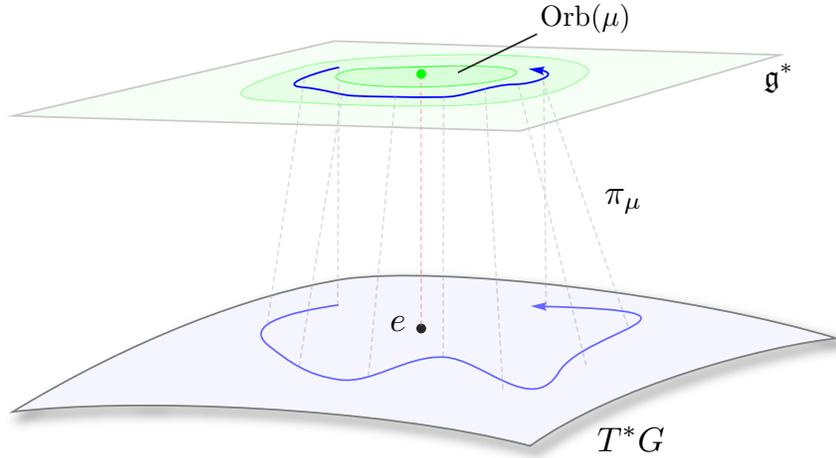
Figure 2.49: For Lie-Poisson systems the reduced dynamics on the coadjoint orbits $\mathcal{O}(\mu)$ on the dual Lie algebra $\mathfrak{g}^*$ are related to the original dynamics on the cotangent bundle T^*G of the group by the quotient projection $\pi_\mu : T^*G \to \mathbf{J}^{-1}(\mu)/G_\mu \cong \mathcal{O}(\mu)$.

But by linearity of the derivative this immediately implies

$$H_\omega = \xi \cdot \omega \tag{2.364}$$

which is the reduced Hamiltonian for the system.

We have seen that for Lie-Poisson systems the reduced phase space is given by the coadjoint orbits $\mathcal{O}_\mu$ on the dual Lie algebra $\mathfrak{g}^*$, and it is a general result that $\mathfrak{g}^*$ decomposes into symplectic leaves formed by such orbits. Hence, it is also possible to discuss Lie-Poisson reduction with a more global perspective where dynamics are directly "dropped" to the dual Lie algebra $\mathfrak{g}^*$.[191] We will briefly discuss this perspective in the following.

Under the left translation action, T^*G/G is diffeomorphic to $\mathfrak{g}^*$. This can be seen by observing that every $(g, \alpha_g) \in T_g^*G$ can be written as the push-forward of some $\alpha \in \mathfrak{g}^* \cong (e, \alpha) \in T_e^*G$,[192] or equivalently as the pullback

$$(g, \alpha_g) \mapsto (e, (T_e^*L_g)\,\alpha_g) = (e, \mathbf{J}_R(\alpha_g)) \tag{2.365}$$

cf. Proposition 2.106 . Hence, the cotangent lift of the left translation action is isomorphic to the action of G on $G \times \mathfrak{g}^*$ given by

$$h \circ (g, \alpha) = (h\,g, \alpha). \tag{2.366}$$

---

[191]In the contemporary literature, Lie-Poisson reduction typically directly proceeds to the dual Lie algebra $\mathfrak{g}^*$ and establish the existence of the Lie-Poisson bracket governing the dynamics. However, we believe that such an approach takes away much of the geometric intuition since the pre-image $\mathbf{J}^{-1}(\mu)$ and the reduced phase space $P_\mu = \mathbf{J}^{-1}/G_\mu = \mathcal{O}(\mu)$ only appear implicitly, and for example also a direct link to Noether's theorem is lost.

[192]More formally, we construct hence a local trivialization of the fiber bundle $\pi : T^*G \to G$.

where the coset of left invariant one forms $(T_e^* L_g) \, \alpha_g$ has been identified with its representative element α in the dual Lie algebra $\mathfrak{g}^*$. But from the above action the quotient $T^*G/G \cong (G \times \mathfrak{g}^*)/G$ is $\mathfrak{g}^*$. With the definition of the Lie-Poisson momentum maps, Proposition 2.106, we can thus consider $\mathbf{J}_R$ as the quotient map π_μ for the reduced phase space $\mathfrak{g}^*$, cf. Def 2.225, that is

$$\mathbf{J}_R : T^*G \to T^*G/G \cong \mathfrak{g}^*. \tag{2.367}$$

where the right momentum map has to be employed by Eq. 2.365. By Proposition 2.107, the momentum maps are Poisson maps, and thus the Poisson reduction theorem in Theorem 2.38 immediately implies that a reduced Poisson structure on $\mathfrak{g}^*$ exists. As one would expect, one can show that the reduced Poisson structure is the Lie-Poisson bracket with the generated dynamics restricted to the symplectic leaves on $\mathfrak{g}^*$, that is to the coadjoint orbits. A schematic overview of the global view of Lie-Poisson reduction is provided in Fig. 2.50. As is apparent from the figure, the most significant difference between Lie-Poisson reduction and symplectic and Poisson reduction is that two momentum maps are employed, a classical one corresponding to Noether's theorem whose image are the conserved quantities in $\mathfrak{g}_-^*$, and a second one which is employed to "drop" the dynamics to the reduced phase space g_+^*.

Remark 2.175. Lie-Poisson reduction is the Hamiltonian counterpart of Euler-Poincaré reduction, and there exists a Legendre transform connecting the two for hyperregular systems.[193]

Remark 2.176. For Lie-Poisson systems a Lie group provides both the configuration space and the system's symmetry group. This makes it sometimes difficult to distinguish what facet of a system one is dealing with and what properties arise from which of the two sides, as is exemplified by the double role played by the dual Lie algebra $\mathfrak{g}^*$, cf. Fig 2.50.

Lagrangian, Eulerian, and Convective Representation[194]   Lie-Poisson systems enable to intrinsically characterize the difference between the Lagrangian, Eulerian, and convective representations of a mechanical system, which one encounters for example in elasticity, fluid dynamics, and the Euler top. The Lagrangian representation is the unreduced one on the cotangent bundle T^*G of the Lie group G, which provides the configuration space for the system, and the Eulerian and convective representations are obtained by right and left translation to the dual Lie algebra $\mathfrak{g}^*$, respectively, see Fig. 2.50. In the following we will briefly connect the abstract ideas of Lie-Poisson systems and momentum maps to more familiar ones such as Lagrangian and Eulerian

---

[193]Holm, Schmah, and Stoica, *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*, Chapter 9.1.

[194]For a more detailed discussion see for example (Marsden and Hughes, *Mathematical Foundations of Elasticity*, Chapter 1).
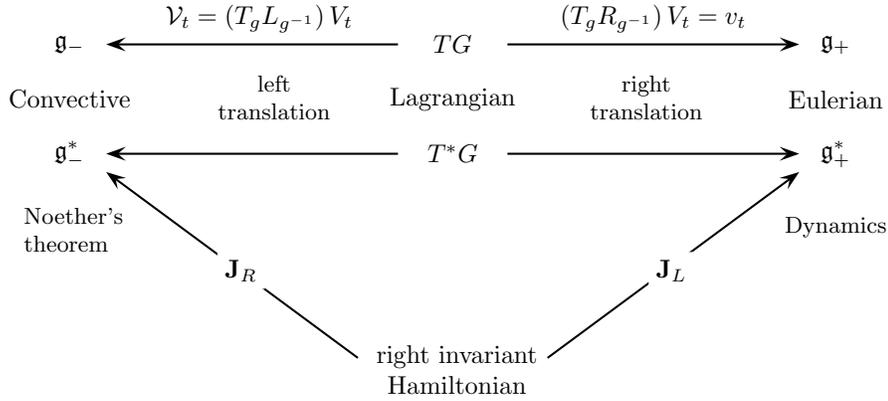
Figure 2.50: Geometry of Lie-Poisson reduction for a right invariant system such as the ideal Euler fluid. For a left invariant Hamiltonian, such as for the Euler top, the lower part of the diagram has to be mirrored, and the conserved quantities are obtained on $\mathfrak{g}_+^*$ while the reduced dynamics are described on $\mathfrak{g}_-^*$.

velocity, and for simplicity we will in the following again assume that we are concerned with a left action.

We recall that the action of a Lie group G on a manifold $\mathcal{M}$ can always be considered using the group homomorphism of G into the diffeomorphisms $\mathrm{Diff}(\mathcal{M})$ of $\mathcal{M}$, which usual yields a subgroup $\mathrm{Diff}_*(\mathcal{M}) \subset \mathrm{Diff}(\mathcal{M})$ representing the effective action of G on $\mathcal{M}$, see Remark 2.142. For example, with $G = \mathrm{SO}(3)$ acting on $\mathbb{R}^3$, the subgroup $\mathrm{Diff}_{\mathrm{rb}}(R^3) \subset \mathrm{Diff}(\mathbb{R}^3)$ of all diffeomorphisms is the set of all orientation preserving isometries of $\mathbb{R}^3$, which are also known as rigid body transformations. The group homomorphism corresponds to choosing a **reference configuration** $\mathcal{M}_0$, which is identified with the identity of the group, and describing all other configurations as deformations or transformation by an element $\varphi \in \mathrm{Diff}_*(\mathcal{M})$, cf. Example 2.146. Hence, $\mathrm{Diff}_*(\mathcal{M})$ can be considered as the configuration space and we obtained a description of the system as Lie-Poisson system. Choosing $\mathcal{M}_0$ to be the configuration at an initial time $t = 0$, time evolution is described by a smooth curve $\varphi_t : [0, \bar{t}] \to \mathrm{Diff}(\mathcal{M})$ in the space of diffeomorphisms $\mathrm{Diff}_*(\mathcal{M})$, and for each "particle" $\tilde{x} \in \mathcal{M}_0$ a trajectory $x(t) = \varphi(\tilde{x}, t) = \varphi_t(\tilde{x})$ in $\mathcal{M}$ is traced out.[195] Using this Lie-Poisson description of the system, the **Lagrangian velocity** is

$$V_t = V(\tilde{x}, t) = \frac{\partial \varphi(\tilde{x}, t)}{\partial t} \in T_{\varphi(\tilde{x}, t)}\mathcal{M} \tag{2.368}$$

where $\tilde{x} \in \mathcal{M}$ is a point in the reference configuration at the initial time, and the velocity V_t is a tangent vector at the current point $x(t) = \varphi(\tilde{x}, t)$ along the trajectory starting at $\tilde{x}$, cf. Fig. 2.51. In contrast, the **Eulerian velocity**

---

[195]In the context of fluid dynamics, the points $\tilde{x}$ in the reference configuration are known as "molecules", although one has to be careful to not interpret this too literal.
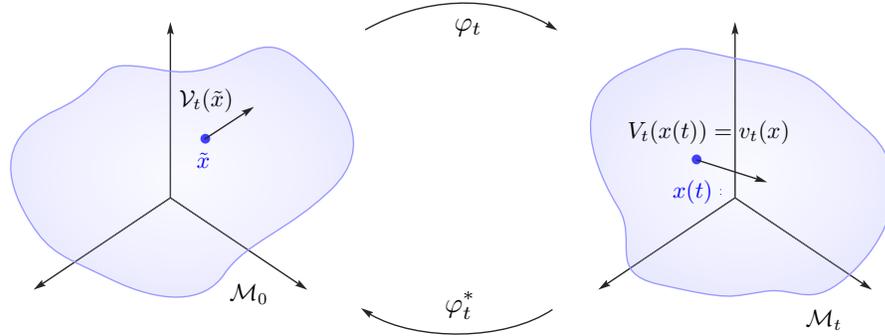
Figure 2.51: The Lagrangian velocity $V_t(\varphi_t(\tilde{x})) = V_t(x(t))$ is the tangent vector to the trajectory $\varphi_t(\tilde{x})$ starting at the point $\tilde{x}$ in the reference configuration at $x(t) = \varphi_t(\tilde{x})$. The Eulerian velocity $v_t(x) = V_t \circ \varphi_t^{-1}$ is the velocity at a fixed location x, while the convective velocity $\mathcal{V}_t(X)$ is the change in $X(t)$ for a fixed x.

$v(x,t)$ is defined for a time invariant location $x \in \mathcal{M}$ and it is given by

$$v_t = v(x,t) = V(\varphi^{-1}(x,t),t) = V_t \circ \varphi_t^{-1} = \dot{\varphi}_t \circ \varphi_t^{-1} \ \in T_x\mathcal{M} \qquad (2.369)$$

where we can think of the inverse diffeomorphism φ^{-1} as choosing $\tilde{x}$ in the reference configuration such that it is at time t under the flow φ_t at the fixed location $x = \varphi(\tilde{x},t)$.[196] The **Eulerian velocity is right invariant** under composition by a (time invariant) diffeomorphism $\psi \in \mathrm{Diff}_*(\mathcal{M})$ which means that[197]

$$v_t \circ \psi = v_t \qquad (2.370)$$

as this important property is shown in the following remark.

Remark 2.177. From the definition of the Eulerian velocity we have

$$\mathrm{v}_t \circ \psi = \frac{\partial}{\partial t}\left(\varphi_t \circ \psi\right) \circ \left(\varphi_t \circ \psi\right)^{-1} = \left(\frac{\partial \varphi_t}{\partial t} \circ \psi\right)\left(\psi^{-1} \circ \varphi_t^{-1}\right) \qquad (2.371a)$$

where the right hand side follows since ψ is time-independent and since for arbitrary maps f, g one has $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. But $\psi \circ \psi^{-1} = \mathrm{id}$ and hence

$$\frac{\partial}{\partial t}\left(\varphi_t \circ \psi\right) \circ \left(\varphi_t \circ \psi\right)^{-1} = \frac{\partial \varphi_t}{\partial t} \circ \varphi_t^{-1} \qquad (2.371b)$$

and the right hand side is the definition of the Eulerian velocity in Eq. 2.369 for φ_t without the right action by ψ. Hence, the Eulerian velocity is indeed right invariant.[198]

---

[196]In the literature, the Eulerian velocity is hence also known as spatial velocity.

[197]Cf. (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, p. 19).

[198]It is easy to see that a composition of φ_t by ψ leads by the chain rule to $\frac{\partial \psi}{\partial x} \frac{\partial x(t)}{\partial t}$ with $x(t) = \varphi(t)$ so that there is no left invariance.

In the right invariance of the Eulerian velocity, we can think of the action of ψ as a redefinition of the particles $\tilde{x}$ in the reference configuration before the flow is started, and the invariance is hence known as **particle relabelling symmetry**. It has to be noted that the right invariance of the Eulerian velocity is a priori unrelated to the invariance of the Hamiltonian, which is required for Noether's theorem to apply, although for a purely kinetic system the two are usually related. The **convective velocity** for a time invariant location $x \in \mathcal{M}$ is the velocity of the trajectory $\tilde{x}(t) = \varphi^{-1}(x, t)$ given by

$$-\mathcal{V}_t = -\mathcal{V}(\tilde{x}, t) = \frac{\partial \tilde{x}(x, t)}{\partial t} = \frac{\partial}{\partial t} \varphi_t^{-1}(x) = \varphi_t^* v_t \in T_{\varphi^{-1}(x)} \mathcal{M} \qquad (2.372)$$

where $\tilde{x}(t) = \varphi^{-1}(x)$ traces out a path in the reference configuration.

The connection between the above definitions of the Eulerian and Lagrangian velocities and the translates to the Lie algebra depicted in Fig. 2.50 is obtained as follows. In Remark 2.177 we showed the right invariance of the Eulerian velocity and by the definition of an invariant vector field on a Lie group, cf. Def 2.177, we therefore have

$$v_t = V_t \circ \varphi_t^{-1} = (T_{\varphi_t} R_{\varphi_t^{-1}}) V_t(\varphi_t) \in T_e G \cong \mathfrak{g}_+ \qquad (2.373)$$

since φ_t^{-1} is acting via tangent lifts on V_t. Hence, right translation of the Lagrangian velocity V_t yields the Eulerian velocity v_t on the right Lie algebra $\mathfrak{g}_+$. For the convective velocity we have by the definition of the pullback for vectors in Def. 2.117 that

$$-\mathcal{V}_t = \varphi_t^* v_t = \varphi_t^* \left(V_t \circ \varphi_t^{-1} \right) = (\varphi_t^{-1})_* V_t = (T\varphi_t^{-1}) V_t \qquad (2.374)$$

where the φ_t^{-1} from the Eulerian velocity $v_t = V_t \circ \varphi_t^{-1}$ is naturally subsumed in the push-forward $(\varphi_t^{-1})_* V_t$ by its definition, cf. Def. 2.117. The question is hence if Eq. 2.374 is a left or a right action, cf. Def. 2.184. For arbitrary $\psi \in \mathrm{Diff}_*(\mathcal{M})$ we have

$$\psi^*(\varphi_t^* v_t) = (\psi^{-1})_* (\varphi_t^{-1})_* v_t \qquad (2.375\mathrm{a})$$

and by Proposition 2.47, i), this equals

$$\psi^* \circ \varphi_t^* v_t = (\psi^{-1} \circ \varphi_t^{-1})_* v_t \qquad (2.375\mathrm{b})$$

which by the definition of a group action shows that it is a left action. Hence, Eq. 2.374 can be written as

$$\mathcal{V}_t = \varphi_t^* v_t = (\varphi_t^{-1})_* V_t = (T\varphi_t^{-1}) V_t = (T_{\varphi_t} L_{\varphi_t^{-1}}) V_t \in T_e G \cong \mathfrak{g}_- \qquad (2.376)$$

and the convective velocity $\mathcal{V}_t$ is obtained by left translation from the Lagrangian velocity V_t, see again Fig. 2.50 and also Fig. 2.51.

	Domain	Lie Algebra	Applications
Diff	$\mathcal{M}$	$(\mathfrak{X}(\mathcal{M}),[\,,])$	general relativity
Diff_μ	$(\mathcal{M},\mu_{\mathrm{vol}})$	$(\mathfrak{X}_{\mathrm{div}}(\mathcal{M}),[\,,])$	ideal Euler fluid, optimal transport
$\mathrm{Diff}_{\mathrm{can}}$	(P,ω)	$(\mathfrak{X}_{\mathrm{Ham}}(P),[\,,])$ $(\mathcal{F}(P),\{\,,\})$	plasma physics, light transport
$\mathrm{Diff}_{\mathrm{con}}$	$(S^*Q,\xi=\sigma^*\theta)$	$(\mathfrak{X}_{\mathrm{Con}}(S^*Q),[\,,])$ $(\mathcal{F}_0(T^*Q),\{\,,\})$	light transport

Table 2.2: Diffeomorphism groups and their applications.

Diffeomorphism Groups as Lie-Poisson Systems We saw before how Lie-Poisson systems arise through a group homomorphism from a group G into a subgroup $\mathrm{Diff}_*(\mathcal{M})$ of the diffeomorphisms of the manifold $\mathcal{M}$ on which G is acting. Many systems in continuum mechanics are Lie-Poisson systems whose configuration space is an infinite-dimensional diffeomorphism group, and which does not arise from a group homomorphism from a finite dimensional group. Such diffeomorphism groups arise because they describe smooth, continuous deformations of a domain which are time reversible—this is the very definition of a diffeomorphism—and for physical reasons they often preserve some additional structure such as the volume form of the symplectic 2-form of a manifold. As briefly already discussed in Chapter 2.3.3.3, intuitively the Lie algebra for a diffeomorphism group can be determined by fixing a single point $\tilde{x} \in \mathcal{M}$ in the domain of the diffeomorphism and considering its time evolution along the curve $x(t) : [a,b] \to \mathcal{M}$ with $x(0) = \tilde{x}$. Differentiating $x(t)$ with respect to time at $t = 0$ yields a vector $\check{X}(m) \in T_m\mathcal{M}$, and extending this to all $m \in \mathcal{M}$ shows that the infinitesimal generator of the flow is a vector field in $T\mathcal{M} \cong \mathfrak{X}(\mathcal{M})$. Hence, the Lie algebra of the diffeomorphism group can be identified with the space of all vector fields $\mathfrak{X}(\mathcal{M})$ on $\mathcal{M}$, and as we saw before the space has a natural Lie bracket given by the Jacobi-Lie bracket for vector fields. When a diffeomorphism preserves some additional structure, then this has to be reflected in the Lie algebra, and for example for the space of volume preserving diffeomorphisms the Lie algebra is given by the space $(\mathfrak{X}_{\mathrm{div}}(\mathcal{M}),[\,,])$ of divergence free vector fields.[199] In Table 2.2, various diffeomorphism groups are summarized, together with their Lie algebras and physical systems they describe.

---

[199]For subgroups of $\mathrm{Diff}(\mathcal{M})$ one obviously has to show that the infinitesimal generators form a subalgebra of $\mathfrak{X}(\mathcal{M})$ and are closed under the Jacobi-Lie bracket, but this is satisfied for the subgroups relevant for physical applications.

Example 2.147.[200] The classical example for a system whose configuration space is a diffeomorphism group is an ideal, incompressible Euler fluid. Let $(\mathcal{M}, g)$ be a Riemannian manifold with g-volume μ, and for simplicity assume $\mathcal{M} \subset \mathbb{R}^n$. An ideal Euler fluid is classically described by the Euler fluid equation,

$$\frac{\partial v_t}{\partial t} + v_t \cdot \nabla v_t = -\nabla p \tag{2.377a}$$

subject to the constraints

$$\nabla \cdot v_t = 0 \tag{2.377b}$$

$$v_t \parallel \partial \mathcal{M} \tag{2.377c}$$

where $v_t = v(x, t) \in \mathfrak{X}(\mathcal{M})$ is the fluid's Eulerian velocity field and $p : \mathcal{M} \to \mathbb{R}$ the fluid's pressure potential, which is uniquely determined by the divergence freeness constraint $\nabla \cdot v_t = 0$. Equivalently, the dynamics of a fluid can also be described by the Helmholtz form of the Euler equation

$$\frac{\partial \omega_t}{\partial t} + \nabla_{v_t} \omega_t = 0 \tag{2.378}$$

where the vorticity vector field $\omega_t = \omega(x, t) \in \mathfrak{X}(\mathcal{M})$ and the velocity v_t are related by

$$\omega_t = \nabla \times v_t. \tag{2.379}$$

For a fixed domain $\mathcal{M}$, the fluid flow along the trajectories defined by v_t can be described by the incompressible flow $\varphi_t : \mathbb{R} \times \mathcal{M} \to \mathcal{M}$ generated by the Eulerian velocity, and since v_t is a smooth vector field these are volume preserving diffeomorphisms of the flow domain $\mathcal{M}$. Hence, with respect to a reference configuration, any state is described by a volume preserving diffeomorphisms $\varphi \in \mathrm{Diff}_\mu(\mathcal{M})$ acting on the left on $\mathcal{M}$, and the infinite dimensional Lie group $\mathrm{Diff}_\mu(\mathcal{M})$ can thus be considered as the configuration space for an ideal fluid. Time evolution is then a curve $\varphi_t : [a, b] \to \mathrm{Diff}_\mu(\mathcal{M})$ on the group $\mathrm{Diff}_\mu(\mathcal{M})$, and in fact it can be shown that it is a geodesic for the metric defined by the fluid Lagrangian

$$L = \int_{\mathcal{M}} \|v_t\|^2 \mu \tag{2.380}$$

and it is easily seen that it is a positive bilinear form as required for a metric. From our general considerations we know that the elements of the Lie algebra

[200]See for example (Arnold and Khesin, *Topological Methods in Hydrodynamics*, Chapter I) and (Marsden and Weinstein, "Coadjoint orbits, vortices, and Clebsch variables for incompressible fluids"; Morrison, "Hamiltonian description of the ideal fluid"). The idea to describe an ideal Euler fluid using volume preserving diffeomorphisms is due to Arnold (Arnold, "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits").

$\mathfrak{g}$ of $\text{Diff}_\mu(\mathcal{M})$ are vector field on $\mathcal{M}$, and since $\text{Diff}_\mu(\mathcal{M})$ is volume preserving these have to be divergence free. It hence follows that

$$\mathfrak{g} = \left(\mathfrak{X}^t_{\text{Div}}(\mathcal{M}), -[\,,]\right) \tag{2.381}$$

and the elements of $\mathfrak{g}$ are the Eulerian velocity fields v_t. Using the Hodge decomposition, cf. Theorem 2.26, it can be shown that the space of divergence free vector fields tangent to the boundary $\mathfrak{X}^t_{\text{Div}}(\mathcal{M})$ is indeed closed under the Jacobi-Lie bracket $[\,,]$.

For infinite dimensional Lie groups, care is required in the definition of the dual Lie algebra $\mathfrak{g}^*$ since it is no longer ensured that the pairing $\langle\,,\rangle : \mathfrak{g} \times \mathfrak{g}^* \to \mathbb{R}$ is non-degenerate, and since in general the dual is no longer unique and different choices for $\mathfrak{g}^*$ lead to different descriptions of the dynamics. For $\mathfrak{g} \cong \mathfrak{X}^t_{\text{Div}}(\mathcal{M})$, the Lie algebra elements are vector fields, and since a vector pairs naturally with a covector, the vector fields pair naturally with covector or 1-form densities $\Omega^1(\mathcal{M}) \times \text{Den}(\mathcal{M})$ with the pairing given by

$$\langle X, A \rangle = \int_{\mathcal{M}} X\,A = \int_{\mathcal{M}} \alpha(X)\,\mu. \tag{2.382}$$

where $X \in \mathfrak{X}^t_{\text{Div}}(\mathcal{M})$ and $A = \alpha\,\mu \in \Omega^1(\mathcal{M}) \times \text{Den}(\mathcal{M})$. However, is the above pairing non-degenerate? By the Leibniz rule for the interior product we have obtain for the right-most integrand $\alpha(X)\,\mu = \alpha(X) \wedge \mu$ that

$$\text{i}_X(\alpha \wedge \mu) = \text{i}_X\alpha \wedge \mu + \alpha \wedge \text{i}_X\mu. \tag{2.383}$$

and the left hand side equals zero since $\alpha \wedge \mu$ is an $(n+1)$-form on the n-manifold $\mathcal{M}$. Hence,

$$\text{i}_X\alpha \wedge \mu = -\alpha \wedge \text{i}_X\mu. \tag{2.384}$$

Writing the integrand $\alpha(X) \wedge \mu$ in Eq. 2.382 as the interior product $\text{i}_X\alpha \wedge \mu$, we obtain with Eq. 2.384 that

$$\langle X, A \rangle = \int_{\mathcal{M}} \text{i}_X\alpha \wedge \mu = -\int_{\mathcal{M}} \alpha \wedge \text{i}_X\mu. \tag{2.385}$$

Let us now assume that α is an exact 1-form, that is $\alpha = \text{d}f$ for some function $f \in \mathcal{F}(\mathcal{M})$. Then, using the Leibniz rule for the exterior derivative, it follows that

$$\text{d}(f \wedge \text{i}_X\mu) = \text{d}f \wedge \text{i}_X\mu + f \wedge \text{d}(\text{i}_X\mu). \tag{2.386}$$

But the vector field X is divergence free. By Cartan's formula and since μ is a volume form we thus have

$$0 = \pounds_X\mu = \text{d}\text{i}_X\mu + \text{i}_X\text{d}\mu = \text{d}\text{i}_X\mu = 0 \tag{2.387}$$

Eq. 2.386 then becomes

$$\text{d}(f \wedge \text{i}_X\mu) = \text{d}f \wedge \text{i}_X\mu \tag{2.388}$$

and we thus obtain for Eq. 2.385 for an exact 1-form $\alpha = \mathrm{d}f$ that

$$\langle X, A \rangle = -\int_{\mathcal{M}} \alpha \wedge \mathrm{i}_X \mu = -\int_{\mathcal{M}} \mathrm{d}(f \wedge \mathrm{i}_X \mu). \tag{2.389}$$

Finally, using Stokes' theorem yields

$$\langle X, A \rangle = -\int_{\partial \mathcal{M}} f \wedge \mathrm{i}_X \mu \tag{2.390}$$

but the integral vanishes since $X \in \mathfrak{X}_{\mathrm{div}}^t(\mathcal{M})$ is tangential to the boundary. We hence have that exact 1-forms $\alpha = \mathrm{d}f$ are "orthogonal" to divergence free vector fields and the pairing $\langle X, \mathrm{d}f \rangle$ is degenerate. For the dual Lie algebra $\mathfrak{g}^*$ we thus have to restrict us to non-exact 1-forms and we obtain

$$\mathfrak{g}^* = \Omega^1(M)/\mathrm{d}\Omega^0(M) \cong \Omega_{\mathrm{ex}}^2(\mathcal{M}) \tag{2.391}$$

where

$$\Omega_{\mathrm{ex}}^2(\mathcal{M}) = \left\{ \omega \in \Omega^2(\mathcal{M}) \mid \omega = \mathrm{d}\alpha \neq 0, \alpha \in \Omega^1(\mathcal{M}) \right\} \tag{2.392}$$

and an element of $\mathfrak{g}^*$ is ever given by the coset $[\alpha] \in \Omega^1(M)/\mathrm{d}\Omega^0(M)$ or by the 2-form $\omega \in \Omega_{\mathrm{ex}}^2(\mathcal{M})$. In fact, the space $\mathfrak{g}^* \cong \Omega_{\mathrm{ex}}^2(\mathcal{M})$ of exact 2-forms can be identified with classical vorticity by

$$\omega = \mathrm{d}v_t^\flat \tag{2.393}$$

where $v_t^\flat \in \Omega^1(\mathcal{M})$ is the 1-form field associated with the Eulerian velocity $v_t \in \mathfrak{X}(\mathcal{M})$ by the musical isomorphism with respect to the metric g on $\mathcal{M}$.[201]

Remark 2.178. It should be noted how a rather careful construction of the dual Lie algebra is necessary in the infinite dimensional case, and respecting this structure is often delicate but unavoidable when the theory is applied.[202]

It can be shown that for the dual Lie algebra $\mathfrak{g}^*$ given by $\Omega_{\mathrm{ex}}^2(\mathcal{M})$ the dynamics on the coadjoint orbits $\mathcal{O}_\mu \subset \mathfrak{g}^*$ are described by the Helmholtz form of the Euler equation in Eq. 2.378. Interestingly, the classical ideal Euler fluid

---

[201]Strictly speaking, this only holds if the first cohomology group of M is trivial, $H_1(M, \mathbb{R}) = 0$, for example if M is simply connected, cf. (Arnold and Khesin, *Topological Methods in Hydrodynamics*, Corollary 7.9).

[202]A point in case is the recent algorithm in (de Witt, "Fluid Simulation in Bases of Laplacian Eigenfunctions"; De Witt, Lessig, and Fiume, "Fluid simulation using Laplacian eigenfunctions"). There the eigenfunctions of the Laplace-Beltrami operator are employed to represent fluid velocity and vorticity. Assuming $\mathcal{M} \subset \mathbb{R}^n$ with the usual metric, the exact 2-forms $\mathrm{d}\alpha$ can be identified with the co-exact 1-forms α_δ using the codifferential operator δ by $\alpha_\delta = \delta\mathrm{d}\alpha = \mathrm{d}\star\mathrm{d}\alpha$ where for the right hand side we used the definition of the codifferential. But $\mathrm{d}\star\mathrm{d}$ is just the Laplace-Beltrami operator when $\alpha = v^\flat$ for a divergence free vector field v. Hence, using the eigenfunctions of the Laplace-Beltrami operator restricts the representable fields to those where velocity and vorticity are in one-to-one correspondence and this avoids in practice the reconstruction of v from ω, for example using a Green's function such as the Biot-Savart law.

equation for the velocity v_t, Eq. 2.377, is obtained when the dual Lie algebra is formed by vector field densities on $\mathcal{M}$, that is $\mathfrak{g}^* = \mathfrak{X}(\mathcal{M}) \otimes \mathrm{Den}(\mathcal{M})$, with the pairing given by

$$\langle X, Y \rangle = \int_{\mathcal{M}} (X \cdot y)\, \mu \tag{2.394}$$

for $X \in \mathfrak{g}$ and $Y = y\,\mu \in \mathfrak{g}^*$.

Since the right invariant Eulerian velocity is employed to describe the dynamics and define the Lagrangian and Hamiltonian for the system, cf. Eq. 2.380, the ideal Euler fluid is a right invariant system.[203] Hence, it follows from the general theory for Lie-Poisson systems that $\mathbf{J}_R$ provides the conserved quantities in the convective representation $\mathfrak{g}^*_-$, and Eq. 2.391 and Eq. 2.393 imply that this is the convective vorticity. Nonetheless, it is instructive to compute the momentum map "bare hands", and we will do so in the following by deriving the "time evolution" equation in the convective representation by pulling back the dynamics from the right to the left Lie algebra.[204] In fact, we already know that we can obtain the convective velocity $\mathcal{V}_t$, defined in the reference configuration $\mathcal{M}_0$, from the Eulerian velocity v_t by the pullback along the flow map φ_t,

$$\mathcal{V}_t = \varphi_t^*\, v_t. \tag{2.395}$$

Additionally, in the following we will also need the pullback of the metric along the flow

$$C_t = \varphi_t^* g \tag{2.396}$$

which, as we already discussed in Example 2.78, is the Cauchy-Green tensor C_t which "measures" the deformation of the time invariant metric g in the current configuration when pulled back into the reference configuration $\mathcal{M}_0$.[205] By the definition of the Lie derivative, the Cauchy-Green tensor has the time evolution equation

$$\frac{\partial C_t}{\partial t} = \frac{\partial}{\partial t}(\varphi_t^*\, v_t) = \varphi_t^* \pounds_{v_t} g = \pounds_{\varphi_t^* v_t} \varphi_t^* g = \pounds_{\mathcal{V}_t} C_t \tag{2.397}$$

and C_t is hence Lie advected in the reference configuration by the convective velocity. With this, we can determine the time evolution equation in the convective representation by pulling back the Eulerian equations. Hence,

$$0 = \varphi_t^* \left(\frac{\partial \omega_t}{\partial t} + \pounds_{v_t} \omega_t \right) \tag{2.398a}$$

and by the transport theorem this equals

$$0 = \frac{d}{dt} \varphi_t^* \omega_t. \tag{2.398b}$$

---

[203]For most systems the left invariance of the Lagrangian also implies the left invariance of the Hamiltonian, and this is obviously given for purely kinetic systems where the Lagrangian and Hamiltonian essentially coincide, see also (Holm, Schmah, and Stoica, *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*).

[204]The derivation was communicated to us by Tudor Ratiu.

[205]Classically, the Cauchy-Green tensor is defined as $C_t = F_t^T F_t$ where the tangent map $F_t = T\,\varphi_t$ is known as deformation "gradient" although it is unrelated to any gradient.

Since exterior derivative and pullback commute we have

$$0 = \frac{d}{dt}\mathrm{d}(\varphi_t^* \, v_t^{\flat_g}).\tag{2.398c}$$

To interpret $\varphi_t^* \, v_t^{\flat_g}$ consider the pairing of this 1-form in the reference configuration with an arbitrary tangent vector X. By the definition of the pullback we then have

$$(\varphi_t^* \, v_t^{\flat_g})(X) = v_t^{\flat_g}((T\varphi_t)X)\tag{2.399a}$$

and by the definition of the musical isomorphism this equals

$$(\varphi_t^* \, v_t^{\flat_g})(X) = g\left(v_t, (T\varphi_t)X\right).\tag{2.399b}$$

Using the identity $(T\eta_t) \circ (T\eta_t^{-1}) = \mathrm{id}$ we obtain

$$(\varphi_t^* \, v_t^{\flat_g})(X) = g\left((T\eta_t) \circ (T\eta_t^{-1})\, v_t, (T\varphi_t)X\right)\tag{2.399c}$$

$$= (\varphi_t^* g)\left((T\eta_t^{-1})\, v_t, X\right)\tag{2.399d}$$

and by the definition of the Cauchy-Green tensor and the convective velocity this yields

$$(\varphi_t^* \, v_t^{\flat_g})(X) = C_t\left(\mathcal{V}_t, X\right).\tag{2.399e}$$

Again using the definition of the musical isomorphism this can be written as

$$(\varphi_t^* \, v_t^{\flat_g})(X) = \mathcal{V}_t^{\flat_{C_t}}(X).\tag{2.399f}$$

Hence, $\varphi_t^* \, v_t^{\flat_g}$ is just the 1-form associated with the convective velocity $\mathcal{V}_t$ for the natural metric $C_t = \varphi_t^* g$ defined in the reference configuration. For Eq. 2.398c we hence obtain

$$0 = \frac{d}{dt}\mathrm{d}\mathcal{V}_t^{\flat_{C_t}}\tag{2.400a}$$

and interpreting $\mathrm{d}\mathcal{V}_t^{\flat_{C_t}}$ as convective voriticity $\Omega_t = \mathrm{d}\mathcal{V}_t^{\flat_{C_t}}$ we have

$$0 = \frac{d}{dt}\Omega_t.\tag{2.400b}$$

Hence, the convective vorticity Ω_t in the reference or body representation is conserved under the time evolution, and it is the image of the right momentum map $\mathbf{J}_R$. Eq. 2.400b can also be interpreted as a modern formulation of Kelvin's circulation theorem since by the change of variables formula we have

$$\int_{S_0} \Omega_t = \int_{S_0} \varphi_t^* \omega = \int_{\varphi_t(S_0)} \omega\tag{2.401}$$

where S_0 is a surface in the convective or reference configuration and $S_t = \varphi_t(S_0)$ its image under the flow, and the circulation around the boundary ∂S_t can be obtained by Stokes' theorem since the vorticity is exact, cf. Fig. 2.52.[206]

---

[206]From the definition of the convective velocity it is not necessarily apparent that it is defined at a fixed location in the convective representation but this follows from the definition of the pullback for vector fields, cf. Def. 2.117. For a precise formulation of the connection between Eq. 2.400b and Kelvin's circulation theorem see (Holm, Marsden, and Ratiu, *The Euler-Poincaré Equations and Semidirect Products with Applications to Continuum Theories*, Theorem 6.2).
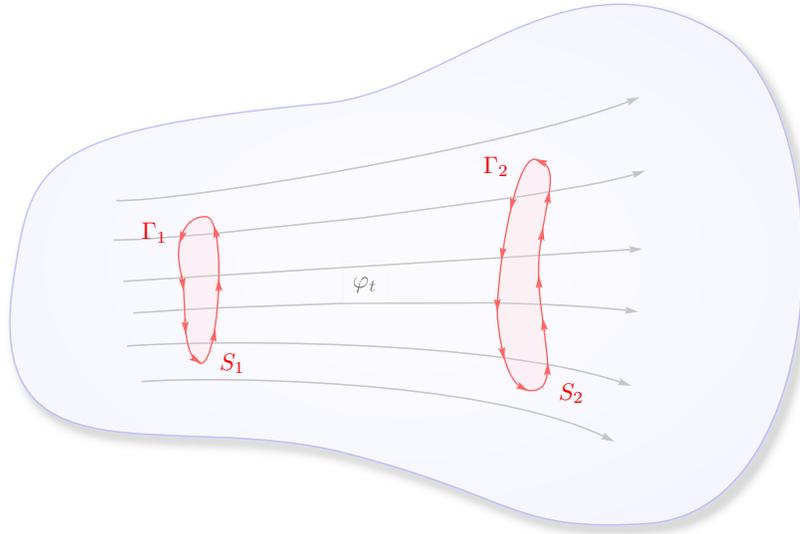
Figure 2.52: Kelvin's circulation theorem states that the circulation $\Gamma_t = \int_{\partial S_t} v_t^\flat$ around the closed loop ∂S_t is conserved along the flow for $S_t = \varphi_t(S_0)$. By Stokes' theorem this is equivalent to the conservation of the integral $\int_{S_t} \omega_t$ of vorticity $\omega_t = \mathrm{d}v_t^\flat$ over the transport surface S_t.

Next to the convective vorticity Ω_t, for the ideal Euler fluid additional conserved quantities exists which are known as enstrophy integrals.[207] For a $2n + 1$ dimensional manifold $\mathcal{M}$ there exists one enstrophy integral

$$J(v_t) = \int_{\mathcal{M}} v_t^\flat \wedge (\mathrm{d}v_t^\flat)^n \tag{2.402a}$$

and for a $2n$ dimensional manifold $\mathcal{M}$ there is an infinite number of integrals

$$J_k(v_t) = \int_{\mathcal{M}} \left(\frac{1}{\mu} (\mathrm{d}v_t^\flat)^n \right)^k \mu \tag{2.402b}$$

for all $k = 1, 2, \ldots$, and in fact even more generally one has

$$J_\phi(v_t) = \int_{\mathcal{M}} \phi \left(\frac{1}{\mu} (\mathrm{d}v_t^\flat)^n \right) \mu \tag{2.402c}$$

for an arbitrary function $\phi : \mathbb{R} \to \mathbb{R}$, and Eq. 2.402b can be considered as a polynomial basis for Eq. 2.402c. Note that the enstrophy integrals are an example of the Casimir functions that were introduced in Chapter 2.3.4.3.[208]

---

[207]For $\mathbb{R}^n$ the result is due to Tartar and Serre ("Invariants et dégénérescence symplectique de l'équation d'Euler des fluides parfaits incompressibles") and for a general Riemannian manifold due to Ovsienko, Khesin, and Chekanov ("Integrals of the Euler equations of multidimensional hydrodynamics and superconductivity").

[208]It has to be noted that although the ideal Euler equation has an infinite number of conserved quantities for an even dimensional, $\mathcal{M}$ it is not an integrable system.

2.3.5.3   Lie-Poisson Systems for Diff_{can}[209]

The diffeomorphism group which will be of interest in the remainder of the thesis is the group $\text{Diff}_{\text{can}}(P)$ of canonical transformations preserving the symplectic 2-form on a symplectic manifold (P, ω). We will consider this group hence in more detail in the following.

Taking the group perspective and studying $\text{Diff}_{\text{can}}(P)$ can be seen as changing perspective from a local point of view, studying single trajectories along the flow of the Hamiltonian vector field, as we did mostly in the foregoing and in particular in Chapter 2.3.4.3, to a global viewpoint where the flow on the entire symplectic manifold is considered. Let us begin by defining the object of interest for this section.

Definition 2.228. *Let (P, ω) be a simply connected, compact symplectic manifold of dimension $2n$. The **group of symplectic diffeomorphisms** or **symplectomorphisms***

$$\text{Diff}_{\text{can}}(P) = \{\eta \in \text{Diff}(P) \mid \eta^* \omega = \omega\}$$

is the set of all diffeomorphisms $\eta \in \text{Diff}(P)$ acting on the left on P that preserve the symplectic 2-form.

When X is the infinitesimal generator of the flow $\eta : \mathbb{R} \times P \to P$, then the above definition can also be stated infinitesimally as $\mathcal{L}_X \omega = 0$. It follows immediately from Def. 2.210 that each element $\eta \in \text{Diff}_{\text{can}}(P)$ is a canonical transformation, and as usual the example to have in mind for a symplectic manifold is the cotangent bundle $(P = T^*Q, \omega = \text{d}\theta)$ where $\theta = p_i dq^i \in \Omega^1(T^*Q)$ is the canonical 1-form and $Q \subset \mathbb{R}^3$.

As explained before for general Lie-Poisson systems, physically we interpret each element $\eta \in \text{Diff}_{\text{can}}$ as a "deformation" of a chosen reference configuration, identified with the identity element of the group, and $\text{Diff}_{\text{can}}(P)$ can then be considered as configuration space. Time evolution is thus described by a smoothly changing family $\eta(t) : [a, b] \to \text{Diff}_{\text{can}}(P)$ of diffeomorphisms and, as usual, the reference configuration is identified with the configuration at the initial time. Physically we are thus interested in the subgroup $\text{Diff}^e_{\text{can}}(P)$ of $\text{Diff}_{\text{can}}(P)$ containing the identity, and for simplicity we will in the following usually identify $\text{Diff}_{\text{can}}(P)$ with $\text{Diff}^e_{\text{can}}(P)$.

Remark 2.179. The relationship between $\text{Diff}^e_{\text{can}}(P)$ and $\text{Diff}_{\text{can}}(P)$ is similar to the situation in Euclidean space $\mathbb{R}^n$ where $O(n)$ describes the set of all transformations preserving the inner product with the rotation group $SO(n)$ being

---

[209]Our discussion is largely based on (Polterovich, *The Geometry of the Group of Symplectic Diffeomorphisms*; McDuff, "Lectures on Groups of Symplectomorphisms"; Gümral, "Geometry of plasma dynamics. I. Group of canonical diffeomorphisms"). More detailed and more mathematical accounts can be found in (McDuff and Salamon, *Introduction to Symplectic Topology*) and interesting recent work on the subject is for example (Ebin, "Geodesics on the Symplectomorphism Group"; Khesin, "Dynamics of symplectic fluids and point vortices"). The functional analytic questions of the infinite dimensional group structure were first investigated in (Ebin and Marsden, "Groups of Diffeomorphisms and the Motion of an Incompressible Fluid").

the connected component of $\mathrm{O}(n)$ containing the identity, and the component not connected to the identity is formed orientation reversing transformations, cf. Example 2.91.

That $\mathrm{Diff}_{\mathrm{can}}(P)$ is closed under composition, and hence forms a subgroup of $\mathrm{Diff}(P)$, is apparent from the definition, and it follows from Liouville's theorem that $\mathrm{Diff}_{\mathrm{can}}(P)$ is also a subgroup of $\mathrm{Diff}_{\mu}(P)$, the group of volume preserving diffeomorphisms.[210] Hence, we have the following inclusions

$$\mathrm{Diff}_{\mathrm{Ham}}(P) \subseteq \mathrm{Diff}^e_{\mathrm{can}}(P) \subseteq \mathrm{Diff}_{\mathrm{can}}(P) \subseteq \mathrm{Diff}_{\mu}(P) \subseteq \mathrm{Diff}(P) \qquad (2.403)$$

and when the inclusions become equalities depends on the domain P. The group $\mathrm{Diff}_{\mathrm{Ham}}(P)$ on the very left of Eq. 2.403 is the group of Hamiltonian diffeomorphisms whose elements are symplectic transformations generated by Hamiltonian vector fields

$$\mathrm{Diff}_{\mathrm{Ham}}(P) = \{\eta \in \mathrm{Diff}_{\mathrm{can}}(P) \mid \xi_P(\eta) = X_H\}. \qquad (2.404)$$

The difference between $\mathrm{Diff}_{\mathrm{Ham}}(P)$ and $\mathrm{Diff}_{\mathrm{can}}(P)$ arises from the topology of P, and since X_H only depends on the exterior derivative $\mathrm{d}H$ of the Hamiltonian H, the discrepancy can be characterized using the de Rahm cohomology $\mathrm{H}^k(P)$ that describes the difference between closed and exact differential k-forms, cf. Def. 2.169. By definition of the Hamiltonian vector field, all vector fields X generating a symplectic flow are Hamiltonian when for each 1-form $\mathrm{i}_X\omega = \omega(X, \cdot)$ there exists globally a function $H \in \mathcal{F}(P)$ such that $\mathrm{i}_X\omega = \omega(X, \cdot) = \mathrm{d}H$. But is equivalent to requiring that the first de Rahm cohomology is trivial,

$$\mathrm{H}^1(P) = 0. \qquad (2.405)$$

We will assume this in the following, that is we will consider the case when $\mathrm{Diff}_{\mathrm{Ham}}(P) = \mathrm{Diff}_{\mathrm{can}}(P)$.

Remark 2.180. The first de Rahm cohomology vanishes when the space is contractible. Hence, the n-ball in $\mathbb{R}^n$, and any solid homeomorphic to the n-ball, has $H^k = 0$, and this also holds for the cotangent bundle over these spaces. A counterexample where $\mathrm{H}^k \neq 0$ is the torus which is clearly not contractible. An advantage of working with $\mathrm{Diff}_{\mathrm{can}}(P)$, rather than $\mathrm{Diff}_{\mathrm{Ham}}(P)$, is that the former has a natural category, those of symplectic manifolds, although these technicalities are beyond the scope of the thesis. For a detailed discussion of the difference between $\mathrm{Diff}_{\mathrm{can}}(P)$ and $\mathrm{Diff}_{\mathrm{Ham}}(P)$ in the general case when $\mathrm{H}^1(\mathcal{M}) \neq 0$ we refer to the literature.[211]

The Eulerian velocity of the action of $\mathrm{Diff}_{\mathrm{can}}(P)$ is the Hamiltonian vector field X_H, since it is the vector field fixed in space whose integral curves defined the flow of the system. Remark 2.177, where we showed that the Eulerian velocity for a left action is right invariant, then implies that also the flow

---

[210] As discussed in Chapter 2.3.3, the question requiring care is functional analytic closure.

[211] Polterovich, *The Geometry of the Group of Symplectic Diffeomorphisms*, Chapter 14.

$\eta_t \in \mathrm{Diff}_{\mathrm{can}}(P)$ and the Hamiltonian H are right invariant. Hence, $\mathrm{Diff}_{\mathrm{can}}(P)$ acting on the right forms a symmetry group, and as a Lie-Poisson system one says it is right invariant. We summarize these properties in the following proposition.

Proposition 2.108. *Let* $\mathrm{Diff}_{\mathrm{can}}(P)$ *act on the symplectic manifold* (P, ω) *from the left. Then* $\mathrm{Diff}_{\mathrm{can}}(P)$ *acting from the right is a symmetry group for the system, and the Hamiltonian vector field* X_H, *the flow* φ_t *generated by* X_H, *and the Hamiltonian H are right invariant.*

A corollary of the above proposition is that Noether's theorem applies for the right action of $\mathrm{Diff}_{\mathrm{can}}(P)$, cf. Theorem 2.35, and that the conserved quantities are obtained in the left dual Lie algebra $\mathfrak{g}^*_-$ corresponding to the convective representation, see again Fig. 2.50.

The right invariance of the Hamiltonian vector field can also be employed to determine the Lie algebra of $\mathrm{Diff}_{\mathrm{can}}(\mathcal{M})$, which, as we know from the general considerations for $\mathrm{Diff}(\mathcal{M})$ before, is formed by vector fields $X \in \mathfrak{X}(P)$ on P. By definition, for any Lie group the Lie algebra $\mathfrak{g}$ is constructed using the left or right invariant vector fields. But since X_H is right invariant, this suggests that the space of Hamiltonian vector fields $\mathfrak{X}_{\mathrm{Ham}}(P)$ forms the Lie algebra $\mathfrak{g}$, and nothing else should be expected when $\mathfrak{g}$ is considered as the infinitesimal action of the group $\mathrm{Diff}_{\mathrm{can}}(P)$. The Lie bracket for $\mathfrak{g}$ is then naturally given by the Jacobi-Lie bracket for vector fields, and we showed before that $\mathfrak{X}_{\mathrm{Ham}}(P)$ is in fact closed under the bracket. With $\mathfrak{g} = (\mathfrak{X}_{\mathrm{Ham}}(P), [\,,\,])$, the Lie algebra homomorphism in Proposition 2.96 shows that an equivalent Lie algebra for $\mathrm{Diff}_{\mathrm{can}}(P)$ is provided by $(\mathcal{F}(P), \{\,,\,\})$, the smooth functions on P under the Poisson bracket. Hence, we have established the following proposition.

Proposition 2.109. *Let* (P, ω) *be a simply connected, compact symplectic manifold. The Lie algebra* $\mathfrak{s}$ *of* $\mathrm{Diff}_{\mathrm{can}}(P)$ *is*

$$\mathfrak{s} = (\mathfrak{X}_{\mathrm{Ham}}, -[\,,\,])$$

where $[\,,\,]$ *is the Jacobi-Lie bracket for vector fields, or equivalently,*

$$\mathfrak{s} = (\mathcal{F}(P), \{\,,\,\})$$

where $\{\,,\,\}$ *is the canonical Poisson bracket on* P.

As we saw in the foregoing, the dual Lie algebra $\mathfrak{s}^*$ plays an outstanding role for Lie-Poisson systems, since both the conserved quantities and the dynamics are defined on the space. For infinite dimensional systems we are however faced with the problem that it is no longer uniquely determined and that it has to be ensured that the pairing $\mathfrak{s} \times \mathfrak{s}^* \to \mathbb{R}$ is non-degenerate. The dual Lie algebra $\mathfrak{s}^*$ for $\mathrm{Diff}_{\mathrm{can}}(P)$ is constructed similar to those for the group of volume preserving diffeomorphisms in Example 2.147, and we will in the following denote the Liouville form, the natural volume form on P, again by $\varpi = \omega^n$.

The Hamiltonian vector fields $X_H \in \mathfrak{X}_{\mathrm{Ham}}(P)$, which form the Lie algebra for $\mathrm{Diff}_{\mathrm{can}}(P)$, naturally pair with 1-form densities $A = \alpha \, \varpi \in \Omega^1(P) \otimes \mathrm{Den}(P)$ with the pairing given by

$$\langle X_H, A \rangle = \int_P \alpha(X_H) \, \varpi. \tag{2.406}$$

We have to determine under which conditions the above pairing is non-degenerate. Considering $\alpha(X_H)$ as the interior product $\mathrm{i}_{X_H} \alpha$, we have by the Leibniz rule

$$\mathrm{i}_{X_H}(\alpha \wedge \varpi) = \mathrm{i}_{X_H} \alpha \wedge \varpi + \alpha \wedge \mathrm{i}_{X_H} \varpi, \tag{2.407}$$

and since the $(2n + 1)$-form $\alpha \wedge \varpi$ on the left hand side vanishes on the $2n$ manifold P, we obtain for Eq. 2.406 that

$$\langle X_H, A \rangle = -\int_P \alpha \wedge \mathrm{i}_{X_H} \varpi. \tag{2.408}$$

Assuming the 1-form α is exact, that $\alpha = \mathrm{d}f$ for some function $f \in \mathcal{F}(P)$, we have by the Leibniz rule that

$$\mathrm{d}(f \wedge \mathrm{i}_{X_H} \varpi) = \mathrm{d}f \wedge \mathrm{i}_{X_H} \varpi + f \wedge \mathrm{d}\mathrm{i}_{X_H} \varpi = \mathrm{d}f \wedge \mathrm{i}_{X_H} \varpi \tag{2.409}$$

where $\mathrm{d}f \wedge \mathrm{i}_{X_H} \varpi$ vanishes since X_H is volume preserving so that $\mathrm{i}_{X_H} \varpi$. Hence,

$$\langle X_H, A \rangle = -\int_P \mathrm{d}(f \wedge \mathrm{i}_{X_H} \varpi) \tag{2.410}$$

and Stokes' theorem immediately implies

$$\langle X_H, A \rangle = -\int_{\partial P} f \wedge \mathrm{i}_{X_H} \varpi \tag{2.411}$$

which vanishes since the flow of X_H preserves the Liouville form. Hence, analogous to fluid dynamics, we have that the dual Lie algebra for $\mathfrak{s} = (\mathfrak{X}_{\mathrm{Ham}}, -[\,,])$ is given by non-exact 1-forms. An expression for the dual Lie algebra for $\mathfrak{s} = (\mathcal{F}(P), \{\,,\})$ can be obtained from Eq. 2.408 when the Liouville form is expanded as $\varpi = \omega^n$. We then obtain

$$\langle X_H, A \rangle = -\int_P \alpha \wedge \mathrm{i}_{X_H} \omega \wedge \omega^{n-1} \tag{2.412}$$

and with the definition of the Hamiltonian vector field this equals

$$\langle X_H, A \rangle = \int_P \mathrm{d}H \wedge \alpha \wedge \omega^{n-1}. \tag{2.413}$$

Applying integration by parts, cf. Example 2.80, and with the boundary terms vanishing on physical grounds, we obtain

$$\langle X_H, A \rangle = \int_P H \wedge \mathrm{d}(\alpha \wedge \omega^{n-1}) = \int_P H \wedge f \tag{2.414}$$

where $f = \mathrm{d}(\alpha \wedge \omega^{n-1}) \in \mathrm{Den}_{\mathrm{ex}}(P)$ is an exact density on P. We hence have shown the following proposition.[212]

Proposition 2.110. *Let (P, ω) be a simply connected, compact symplectic manifold, and $\mathrm{Diff}_{\mathrm{can}}(P)$ the group of symplectic diffeomorphisms of P. When the Lie algebra of $\mathrm{Diff}_{\mathrm{can}}(P)$ is $\mathfrak{s} = (\mathfrak{X}_{\mathrm{Ham}}, -[\,,\,])$, then the dual Lie algebra is*

$$\mathfrak{s}^* = \Omega^1(P)/\mathrm{d}\Omega^0(P) \cong \Omega^2_{\mathrm{ex}}(P),$$

and when $\mathfrak{s} = (\mathcal{F}(P), \{\,,\,\})$, then the dual Lie algebra is

$$\mathfrak{s}^* = \mathrm{Den}_{\mathrm{ex}}(P).$$

Remark 2.181. Following the literature, we will in the following usually identify $\mathrm{Den}_{\mathrm{ex}}(P)$ with $\mathrm{Den}(P)$.

Given the dual Lie algebra, we obtain from the general theory of Lie-Poisson systems immediately the Lie-Poisson bracket for $\mathrm{Diff}_{\mathrm{can}}(P)$ on $\mathfrak{s}^*$,

$$\{F, G\}(f) = \int_P f \left\{ \frac{\delta F}{\delta f}, \frac{\delta G}{\delta f} \right\}_P \varpi \tag{2.415}$$

where $f \in \mathfrak{s}^* \cong \mathrm{Den}(P)$, $F, G \in \mathcal{F}(\mathfrak{s}^*)$, and the pairing is indeed well defined since $\{\,,\,\}_P : \mathfrak{s} \times \mathfrak{s} \to \mathfrak{s} \cong \mathcal{F}(P)$ is the canonical Poisson bracket on P.[213] For the reduced Hamiltonian vector field X_H^μ for $\mathrm{Diff}_{\mathrm{can}}(P)$, which by the general theory is given by

$$\dot{\mu} = X_H = -\mathrm{ad}^*_{\frac{\delta H}{\delta \mu}}(\mu), \tag{2.416}$$

cf. Theorem 2.40 and with an additional minus sign since we consider a right invariant system, we have to determine the infinitesimal generator $\mathrm{ad}^*_\xi \mu$ of the coadjoint action. With $\mathfrak{s} = (\mathcal{F}(P), \{\,,\,\})$ and $\mathfrak{s}^* = \mathrm{Den}(P)$, we have by definition of the ad^* operator that

$$\langle \mathrm{ad}^*_G(f), H \rangle = \langle f, \mathrm{ad}_G(H) \rangle \tag{2.417a}$$

and since $\mathrm{ad}_\xi(\eta) = [\xi, \eta]$ and for the chosen Lie algebra we obtain

$$\langle \mathrm{ad}^*_G(f), H \rangle = \langle f, \{G, H\} \rangle. \tag{2.417b}$$

The pairing between $\mathfrak{s}$ and $\mathfrak{s}^*$ is given by integration so that

$$\langle \mathrm{ad}^*_G(f), H \rangle = \int_P f \, \{G, H\} \, \varpi \tag{2.417c}$$

---

[212]For discussions of the dual Lie algebra of $\mathrm{Diff}_{\mathrm{can}}$ see also (Gümral, "Geometry of plasma dynamics. I. Group of canonical diffeomorphisms", Section I.B) and (Khesin, "Dynamics of symplectic fluids and point vortices"), and our construction largely follows these authors.

[213]An explicit derivation of this bracket can be found in (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems – Online Supplement*, p. 76).

and with Corollary 2.21 and identifying functions and densities we obtain

$$\langle \text{ad}_G^*(f), H \rangle = \int_P \{f, G\} \, H \varpi \qquad (2.417\text{d})$$

This shows that

$$\text{ad}_G^*(f) = \{f, G\} = -\{G, f\} \qquad (2.418)$$

and we hence established the following proposition.

Proposition 2.111. *Let (P, ω) be a simply connected, compact symplectic manifold, and $\text{Diff}_{\text{can}}(P)$ the group of symplectic diffeomorphisms of P with Lie algebra $\mathfrak{s} = (\mathcal{F}(P), \{,\})$ and dual Lie algebra $\mathfrak{s}^* = \text{Den}(P)$. Then the reduced Hamiltonian vector field $X_H \in T\mathfrak{s}^*$ is*

$$X_H = \dot{f} = -\text{ad}_{\frac{\delta H}{\delta f}}^*(f) = -\left\{f, \frac{\delta H}{\delta f}\right\}$$

and the time evolution on the coadjoint orbits of $\mathfrak{s}^*$ is described by minus the canonical Poisson bracket on P.

Remark 2.182. For the volume preserving diffeomorphism groups with $\xi \in \mathfrak{g}$ a divergence free vector field the infinitesimal generator of the coadjoint action is given by[214]

$$\text{ad}_\xi^*(\eta) = -\pounds_\xi \, \eta. \qquad (2.419)$$

The result also has to apply for $\text{Diff}_{\text{can}}(P)$ since it is a subgroup of $\text{Diff}_\mu(P)$. In fact, using Proposition 2.94 we obtain for the Hamiltonian vector field in Proposition 2.111 for $\xi \in \mathfrak{s}$ that

$$\text{ad}_\xi^*(f) = \{f, \xi\} = \xi[f] \qquad (2.420\text{a})$$

and with the definition of the Lie derivative for a function in Def. 2.154 we have

$$\text{ad}_\xi^*(f) = \pounds_\xi f. \qquad (2.420\text{b})$$

which coincides with Eq. 2.419 for the chosen Lie algebra for $\text{Diff}_{\text{can}}(P)$.

Remark 2.183. Canonical transformations and their formal group structure are known to physicists for a long time.[215] Emphasizing the group picture means to take the vantage point of modern physics where symmetries are central. This enables for example to see parallels that are not apparent when a system is considered locally, and it also opens up an entirely new set of mathematical tools that can be employed to study a system.

---

[214]See Khesin ("The Group and Hamiltonian Description of Hydrodynamical Systems", Proposition 5.1) and Arnold and Khesin (*Topological Methods in Hydrodynamics*, Theorem 7.5, Corollary 7.6).

[215]See for example the classic text by Goldstein (*Classical Mechanics*, Chapter 9).

Remark 2.184. Van Hove[216] showed that the group associated with classical mechanics is, strictly speaking, not the group of canonical transformation $\mathrm{Diff}_{\mathrm{can}}(P)$ but transformations of the $2n + 1$ dimensional contact manifold $\mathbb{R} \times P$ that leave invariant the Pfaffian

$$\bar{w} = ds - \sum_i^n p_i dq^i \ \in \Omega^1(\mathbb{R} \times P). \tag{2.421}$$

From a physical point of view, the Pfaffian $\bar{w}$ is the Lagrangian, related to the Hamiltonian $H(q, p)$ by

$$\bar{\omega} = \sum_i^n p_i dq^i - H(q, p)dt, \tag{2.422}$$

and it is a result by Pfaff that $\bar{\omega}$ can be written in the canonical form in Eq. 2.421. Since the Lagrangian can be considered as the generator of the dynamics, and since canonically the Hamiltonian is obtained from it using the Legendre transform, it is at least intuitively clear that the group leaving invariant Eq. 2.422 and Eq. 2.421 is an appropriate setting for classical mechanics. $\mathrm{Diff}_{\mathrm{can}}(P)$, which, as we saw before, is the group usually associated with classical mechanics, is the quotient Γ/C where C is the center of Γ,[217] which correspond to translations in s, that is time translations. One application where the difference between Γ and the quotient $\Gamma/C \cong \mathrm{Diff}_{\mathrm{can}}(P)$ becomes important are unitary representations, cf. Section 2.3.3.3 and in particular Remark 2.147. $\Gamma/C = \mathrm{Diff}_{\mathrm{can}}(P)$ leads to a single representation as shown by Koopman, Stone, and von Neumann, cf. Remark 2.129, while Γ has an infinite sequence of representations.[218] Much more recently, it was also shown that the distinction between Γ and the quotient Γ/C is important when finite sub-algebras of $(C^\infty, \{\,,\})$ are sought, which is for example of importance for numerical computations.[219]

---

[216]Van Hove's thesis work, where most of the results first appeared, has recently been made available in an English translation (*On Certain Unitary Representations of an Infinite Group of Transformations*); a summary was was published as (van Hove, "Sur le probléme des relations entre les transformations unitaires de la Mécanique quantique et les transformations canoniques de la Mécanique classique.").

[217]Recall that the center $C(G)$ of a group G is $C(G) = \{g \in G \mid gh = hg, h \in G\}$, that is the set of all elements in G that commute with all other elements in the group.

[218]Van Hove's work on the representation theory laid the foundation for what today is called geometric quantization, and which play an important role in the sequel in Chapter 3.

[219]Scovel and Weinstein, "Finite Dimensional Lie-Poisson Approximations to Vlasov-Poisson Equations".

Chapter 3

The Geometry
of Light Transport

"Books on physics are full
of complicated mathematical formulae.
But thought and ideas, not formulae,
are the beginning of every physical theory."[1]

Let us return to 1939 and the preface of Gershun's book, where it continues: "Valuable as the methods [proposed by Gershun] are, however, they probably do not constitute the ultimate solution of the problem. The light field considered in this book is a classical three-dimensional vector field. But the physically important quantity is actually the *illumination*, which is a function of *five* independent variables, not three. Is it not possible that a more satisfactory theory of the light field could be evolved by use of modern tensor methods in a five-dimensional manifold?"[2] After more than 70 years, we will provide an affirmative answer to this question in the present section.

In Chapter 3.2.1 we will employ the Wigner transform to "lift" Maxwell's equations from configuration to phase space. There, light transport theory is obtained as the short wavelength limit of the phase space description. The energy of the electromagnetic field is then represented by the phase space light energy density—a generalized radiance—and its time evolution in media with possibly spatially varying refractive index is described as a canonical Hamiltonian system. The light transport equation will be obtained from the Hamiltonian description in Chapter 3.2.2, and there we will also study the change of variables when the fibers of phase space are parametrized in spherical coordinates. The connection of light transport theory to Fermat's principle will be established in Chapter 3.2.3 using a Legendre transform, and this will also lead to the natural metric for the theory and that the phase space light energy

---

[1] Albert Einstein and Leopold Infeld, *The Evolution of Physics*, 1938.

[2] Again from Moon and Timoshenko's preface to Gershun's book ("The Light Field"); italics in the original.

density is transported along geodesics for the metric. In Chapter 3.2.4 we will show that the dynamics for light transport can be "dropped" from the cotangent bundle to the cosphere bundle—and this will provide the precise answer to the above question on the "use of modern tensor methods in a five-dimensional manifold" which can be misleading when interpreted too naïvely. The phase space light energy density will be studied in more detail in Chapter 3.2.5, and it will be related to classical concepts such as radiance and irradiance in Chapter 3.2.6 where we consider measurements. After a discussion of these results in Chapter 3.2.8, we will study in Chapter 3.3 the symmetry group of light transport in an idealized setting and show that it is a Lie-Poisson system for the group of canonical transformations. A consequence of this structure is that the conservation of the phase space light energy density, generalized radiance, along a ray trajectory is equivalent to Kelvin's circulation theorem in ideal fluid dynamics, and this will be formally established in Chapter 3.3.3. A unitary representation for the group structure of ideal light transport will be established in Chapter 3.3.4, and this will connect our results to the operator formulation that has been employed previously for light transport.

However, before we consider technical questions, we will in Chapter 3.1 discuss how light transport theory developed since the 18th century. The historical context will be important in the following to interpret the existing theory from a modern point of view, and to establish the connection between classical radiometry and our results.[3]

3.1   A Short History of Light Transport Theory[4]

Early History   The transport of light intensity was already considered in the Middle Ages, both in the Arab World, for example by Ibn al-Haytham, and in Europe, for example by Roger Bacon, John Pecham, and Witelo. In contrast to ancient times, intensity was understood as an objective property of light rays, independent of an observer, and one distinguished between specular and diffuse reflection. In the 16th century, Francesco Maurolico associated light intensity with the density of rays, an idea which can be found to this date,[5] and

---

[3]Mach formulated the importance of understanding the historical development of a scientific concept as follows: "Jeder, der den ganzen Verlauf der wissenschaftlichen Entwicklung kennt, wird natürlich viel freier und richtiger über die Bedeutung einer gegenwärtigen wissenschaftlichen Bewegung denken als derjenige, welcher, in seinem Urteil auf das von ihm selbst durchlebte Zeitelement beschrÃďnkt, nur die augenblickliche Bewegungsrichtung wahrnimmt.", (Mach, *Die Mechanik in ihrer Entwicklung: historisch-kritisch dargestellt*, p. 7); [Everybody, who knows the development of the sciences, will surely think more clearly and correctly about the importance of a contemporary scientific movement than somebody, restricted in his judgement by the time he himself experienced, who can only perceives the current directions.], the authors translation.

[4]The material in this section is largely drawn from the extensive introduction to DiLaura's translation of Lambert's work *Photometry or On The Measure and Gradation of Light, Colors, and Shade* which is often referred to as *Photometria*, the main part of its Latin title. The reproductions in this section are also from this book.

[5]The idea to associate the density of rays with the intensity has a striking similarity with the "honey comb" picture propagated as an intuitive model for differential 2-forms by Misner,

in the early 17th century Johannes Kepler was the first to distinguish between rays, along which intensity is constant, and pencils of ray, for which intensity decays quadratically. Around the same time, Marin Mersenne stated the cosine law, and he also did much experimental work where he clearly distinguish illumination power and brightness, that is physical and psychological aspects.

18th Century In the late 17th and early 18th century, Christiaan Huygens and Anders Celsius were the first to compare the intensity of starlight, for example to determine the distance of stars. An early master of radiometry was in the mid-18th century Pierre Bouguer, whose results appeared in his *Essai d'Optique* and *Traité d'Optique*.[6] He studied for example the decay of light intensity in the atmosphere, which led him to the Beer-Lambert law, and he also performed much experimental work using measurement apparatuses designed by him on the insight that the human eye cannot accurately determine absolute intensity but judge well the relative brightness difference.[7]

Hence, in his experiments light intensity from two different sources was presented next to each other on a surface and a human observer had to judge the relative difference, see the figure to the right and Fig. 3.1. Bouguer also investigated in detail light reflection from rough surfaces and he introduced the idea of modeling them by small, randomly oriented mirrors whose directional distribution determines the macroscopic reflectional properties. However, he was not able to quantify
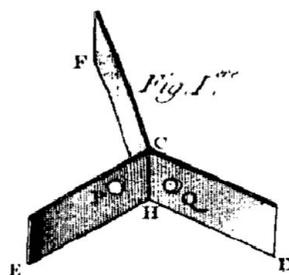
his ideas, and for example determine reflectance values from "mirror" arrangements. In the mid 18th century, Leonhard Euler calculated the incident flux from stars and planets, which necessitated to integrate over a solid angle and required his exquisite command of calculus, and he was hence the first to successfully treated area light sources. Around the same time,[8] Johann Heinrich Lambert became the second classical master of radiometry, and his work can be considered as the climax and synthesis of the ideas of his predecessors. Lambert's experimental work relied on two premises: firstly, the inability of the human eye, the only available measurement device at the time, to objectively and quantitatively determine the difference between two unequal illumination fields, slightly refining the earlier notions of Bouguer, and, secondly, the certainty of the inverse square law and the cosine law. Lambert thus devised experimental setups that establish equality of brightness, which the human eye can judge accurately, and then derived the sought quantities, such as reflectance, trans-

---

Thorne, and Wheeler (*Gravitation*).

[6]Bouguer, *Essai d'optique sur la gradation de la lumiere*; Bouguer, *Traité d'Optique sur la gradation de la lumière*.

[7]This is despite the fact that the human eye can perceive extremely low energies close to single "photons", cf. (Hecht, Shlaer, and Pirenne, "Energy, Quanta, and Vision").

[8]Incidentally, Euler and Lambert were for a short period, from 1764 to 1766, both at the Prussian Academy of Sciences in Berlin, and it was Euler who brought Lambert to Berlin.

mittance, and absorbtance, from the geometry of the experiment. Lambert's contribution to radiometry, however, was not his experimental work but the systematic theoretical framework for the quantification of light intensity that he developed in his *Photometria*.[9] In contrast to most of his contemporaries which still largely relied on geometry, he thereby built on Euler's work and employed infinitesimal calculus to quantify light intensity.[10] Despite the theoretical character, Lambert's *Photometrica* was entrenched in his experimental work which he employed to verify calculations. He therefore also avoided any discussion of the cause of the observations, and he made clear that he was concerned with phenomenological aspects of light intensity that admitted verification by experiment.[11] For the concept today known as radiance, Lambert used the term 'splendor', and 'brightness' referred in his work to surface flux or irradiance,[12] although it is important to keep in mind that, because of his experimental work, for Lambert all quantities were directly related to measurements and not to transport processes. Arguably the most important individual results in *Photometria* are a clear enunciation of the cosine law,[13] which today is named in Lambert's honour, and the linearity of light intensity, and he also distinguished between flux and densities, the flux being constant while the density has a dependence on the refractive index.

The work by Bouguer and Lambert in the 18th century was not only the culmination of earlier efforts, but also the branching point of radiometry from geometric optics. Until then, intensity propagation was considered as part of optics, but with Bouguer's *Essai d'Optique* and *Traité d'Optique* and Lambert's *Photmetria* which were the first books solely concerned with radiometry, it became a research field in its own.

19th Century In the 19th century, the intensity of light was mainly investigated within a larger context, such as the newly emerging fields of black body radiation and astrophysics, and it served then usually as a tool, despite some contribution that were made to the theory itself.[14] For example, the quadratic

---

[9]Lambert and DiLaura, *Photometry or On The Measure and Gradation of Light, Colors, and Shade*.

[10]Lambert heavily employed mathematics in his *Photometria*, which is more in the spirit of Lagrange's *Mécanique analytique*, appearing 50 years later, than of contemporary scientific books, and this has been made responsible for the little attention the work received at its time (Lambert and DiLaura, *Photometry or On The Measure and Gradation of Light, Colors, and Shade*, p. xcvi).

[11]At Lambert's time, still Newton's corpuscular theory of light was the prevailing scientific model, and wave theory only emerged in the first half of the 19th century.

[12]Since Lambert's only measurement device was the human eye, one can in fact argue that splendor and brightness refer to the photometric concepts of luminance and luminous intensity (private communication with D. L. DiLaura). It should be noted that Lambert wrote in Latin and the terminology is hence already subject to the interpretation of the translator.

[13]Classically, Lambert's law refers to the cosine decay of the intensity at a diffuse light source. In the computer graphics literature, however, the term is often used to referred to the foreshortening factor at a receiving surface, see for example (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*).

[14]See Planck's booklet (*Theorie der Wärmestrahlung*) for a summary of the results on black body radiation and how radiometry was employed in this work.
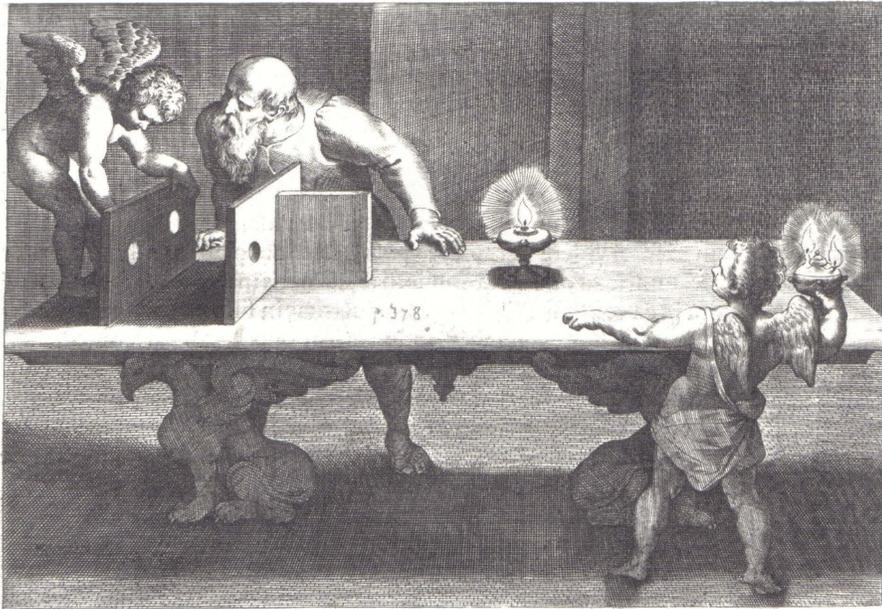
Figure 3.1: Rendering of a radiometric measurement by Pieter Paul Rubens.[†]

dependence of light intensity on the refractive index was first shown by Kirchhoff and Clausius in their work on black bodies,[15] and von Helmholtz obtained the same result in research on the theoretical limits of optical microscopes.[16] Work on the subject was also done by other leading researchers of the time, such as von Humboldt and Fraunhofer,[17] although none of them revisited the physical and mathematical foundations of radiometry. In astronomy, radiometric techniques were in particular employed for the classification of the brightness of stars, for example by Herschel, and there were extensive research efforts in the United States and Germany in the late 19th century for a systematic classification. Radiative transfer theory, where in contrast to radiometry the propagation of the intensity is of central interest and no longer measurements within simple geometric arrangements, was developed in the late 19th century in work by Chwolson[18] and Lommel.[19]. The problem was also of much importance

---

[15]Kirchhoff, "Ueber den Zusammenhang zwischen Emission und Absorption von Licht und Waerme"; Clausius, "Ueber die Concentration von Wärme- und Lichtstrahlen und die Gränzen ihrer Wirkung".

[16]Helmholtz, "Die theoretische Grenze für die Leistungsfähigkeit der Mikroskope".

[17]Many of the results were published in Poggendorff's Annalen, the preeminent publication venue at the time. The journal was officially named 'Annalen der Physik', although it was commonly referred to as 'Poggendorff's Annalen' after its editor in chief at the time (and for more than 50 years).

[†]Reproduced from Lambert and DiLaura, *Photometry or On The Measure and Gradation of Light, Colors, and Shade*, p. liii.

[18]Chwolson, "Grundzüge einer mathematischen Theorie der inneren Diffusion des Lichtes".

[19]Lommel, "Die Photometrie der diffusen Zurückwerfung".

for the young field of astrophysics,[20] and similar results were obtained there somewhat later by Schwarzschild[21] and Schuster.[22] The new radiative transfer models accounted for volume scattering, beyond the simple attenuation that could be modelled using the Beer-Lambert law,[23] and the equilibrium between incident and exitant radiation was considered. Although radiative transfer employs many radiometric concepts and is best understood as an extension of the theory, Lambert's contribution to the subject were largely forgotten in the 19th century, and many of the results in *Photometria* were then only available as "folk knowledge".[24]

20th Century In the 19th century, radiometry still received the attention of the leading scientists at the time, at least within the investigation of other problems. Similar to the situation for the rest of mechanics, see our introductory remarks to Chapter 2.3, this changed with the emergence of the "new" physics in the first decades of the 20th century. Then, radiometry and radiative transfer were useful tools only in certain applications where the "old" physics was still needed, cementing the "arrested development" that was diagnosed by Moon and Timoshenko.[25] Moreover, through many new applications that arose, such as illumination engineering, remote sensing, climate research, optics, radiation hydrodynamics, and computer graphics, radiometry was fragmented and ceased to exist as a coherent research field. As a consequence, terminology diverged between different communities, and concepts and ideas became reinvented in various fields,[26] while the fabric of the theory received no attention, since it was considered to be of no primary importance for the respective applications. This for example also applies to the much cited work by Chandrasekhar,[27] which much refined the existing theory, but did not question its structure. A notable exception is the work by Gershun[28] who employed vector calculus to

---

[20]We follow Mishchenko who argues that Chwolson and Lommel and not Schwarzschild and Schuster are to be credited, cf. Mishchenko, "Radiative Transfer: A New Look of the Old Theory".

[21]Schwarzschild, "Über das Gleichgewicht der Sonnenatmosphäre", reprinted in Schwarzschild, *Gesammelte Werke (Collected Works)*.

[22]Schuster, "Radiation Through a Foggy Atmosphere".

[23]The contribution of Beer on the subject was that he related the extinction coefficient to chemical parameters.

[24]Arguably, a reason for the neglect of Lambert's work was its Latin language, which was no longer the *lingua franca* of scientific discourse in the 19th century. The situation only changed at the end of the 19th century when a German translation of parts of *Photometria* became available (Lambert and Anding, *Photometrie*).

[25]Preface to (Gershun, "The Light Field").

[26]For different accounts of the theory see for example Moon, *The Scientific Basis of Illumination Engineering*; Mihalas and Weibel Mihalas, *Foundations of Radiation Hydrodynamics*; Thomas and Stamnes, *Radiative Transfer in the Atmosphere and Ocean*; Peraiah, *An Introduction to Radiative Transfer: Methods and Applications in Astrophysics*; Castor, *Radiation Hydrodynamics*; Viskanta, *Radiative Transfer of Combustion Systems: Fundamentals and Applications*; Schaepman-Strub et al., "Reflectance Quantities in Optical Remote Sensing –Definitions and Case Studies".

[27]Chandrasekhar, *Radiative Transfer*.

[28]Gershun, "The Light Field".

reformulate radiometry. However, radiometry and radiative transfer are defined on a five dimensional space, and hence beyond the confinements of classical vector calculus.

The Historical Legacy of Radiometry Radiometry was developed by Lambert, based on the work of his predecessors and especially Bouguer, in the middle of the 18th century. The 19th century saw the use of radiometry in other fields, by such eminent scientists as Helmholtz, Kirchhoff, Clausius, Schwarzschild, von Humboldt, and Fraunhofer, and the extension to radiative transfer theory, but without significant changes to Lambert's legacy. Many more applications arose in the 20th century, which led to a fragmentation of the field, although the formulation remained again largely untouched.

Historically, it is remarkable that a theory formulated around 1760 remained almost unchanged for more than 250 years. The basic postulates in Lambert's *Photometria*, such as the inverse square law, the cosine law, or linear superposition of intensity, are still in use today, and even the arguments that are employed to establish them have been altered little since then. Nonetheless, much changed in the rest of the sciences and mathematics since Lambert devised his work, more than a century before Maxwell's equations were discovered and vector calculus was adopted, and long before tensor calculus and its importance to physics became known. Hence, one might question, as did Moon and Timoshenko already again more than 70 years ago,[29] if the formulation is still mathematically the most insightful and vigorous one? Similarly, although Lambert made clear that his theory is purely phenomenological, one might ask today how it is related to the modern understanding of the nature of light and electromagnetic waves, and if measurements of electromagnetic radiation, under suitable conditions paralleling those that led Lambert to his theoretical results, are indeed well approximated by the formulas he devised theoretically and established experimentally?

A Comparison: A Short History of Fluid Dynamics[30] As a counterpoint to the "arrested development" encountered for radiative transfer we briefly want to contrast it with the history of fluid dynamics, which also provides evidence for the significant practical benefits that can be obtained with a modern mathematical formulation. As is well known, the equations of motion for an ideal fluid were first found by Euler who employed basic transport assumptions for the derivation. In the mid 19th century, Clebsch[31] showed that it is possible to derive Euler's equations using Hamilton's form of mechanics, although he had to employ certain non-physical potentials for the formalism to apply. At about the same time, von Helmholtz introduced an alternative perspective on

---

[29]Preface to (ibid.).

[30]Our discussion of the early history is largely drawn from a paper by Christodoulou ("The Euler Equations of Compressible Fluid Flow").

[31]Clebsch, "Über eine allgemeine Transformation der hydrodynamischen Gleichungen"; Clebsch, "Ueber die Integration der hydrodynamischen Gleichungen."

fluid dynamics based on vorticity, using point vorticies, vortex filaments, and vortex sheets, and, building on von Helmholtz's work, Lord Kelvin formulated the circulation theorem which now bears his name. Poincaré was the first to consider ideal fluid dynamics from a semi-geometric perspective. In the early 20th century, he showed that the equations of motion for an ideal fluid, as well as those for a rigid body and the heavy top, can be derived from a variational principle on a Lie algebra,[32] yielding what Marsden and Scheurle called almost a century later the Euler-Poincaré equations.[33] Poincaré's ideas received however only little attention at the time, and Hamel[34] and Ehrenfest[35] seem to have been the only ones to recognize their importance. This only changed in the mid 1960s, when Arnold introduced the modern formulation of ideal hydrodynamics as a geodesic flow on the group $\mathrm{Diff}_\mu(\mathcal{M})$ of volume preserving diffeomorphisms.[36] Arnold also demonstrated how the system admits a canonical Lagrangian and Hamiltonian formulation, without Clebsch's unphysical potentials, by considering the group itself as phase space. Shortly after Arnold's work, Ebin and Marsden put his ideas on a firm mathematical basis and studied the functional analytic questions which arise in the context of infinite dimensional Lie groups.[37] The geometric nature of Clebsch's potentials and von Helmholtz's vortex primitives was clarified by Marsden and Weinstein and they also showed that Kelvin's circulation theorem has a natural geometric interpretation as a generalized Noether theorem.[38] Next to consolidating the existing theory and providing a unified perspective, the modern geometric formulation of ideal fluid dynamics led to important new results, such as estimates on the stability of fluid flow,[39] improved existence and uniqueness results,[40] novel conditions and properties of steady flows,[41] characterizations of control and

---

[32]Poincaré, "Sur une forme nouvelle des équations de la méchanique".

[33]Marsden and Scheurle, "Lagrangian reduction and the double spherical pendulum"; Marsden and Scheurle, "The Reduced Euler-Lagrange Equations".

[34]Hamel, "Die Lagrange-Eulerschen Gleichungen der Mechanik"; Hamel, *Theoretische Mechanik*.

[35]See the discussion in Klein's biography (*Paul Ehrenfest*) of Ehrenfest.

[36]Arnold, "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits"; Arnold, "Sur un principe variationnel pour les découlements stationaires des liquides parfaits et ses applications aux problèmes de stabilité non linéaires"; Arnold, "Hamiltonian character of the Euler equations of the dynamics of solids and of an ideal fluid".

[37]Ebin and Marsden, "Groups of Diffeomorphisms and the Motion of an Incompressible Fluid".

[38]Marsden and Weinstein, "Coadjoint orbits, vortices, and Clebsch variables for incompressible fluids".

[39]Arnold, "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits"; Arnold, "Sur un principe variationnel pour les découlements stationaires des liquides parfaits et ses applications aux problèmes de stabilité non linéaires".

[40]Ebin and Marsden, "Groups of Diffeomorphisms and the Motion of an Incompressible Fluid".

[41]Arnold and Khesin, *Topological Methods in Hydrodynamics*.

controllability of fluid systems,[42] estimates on shock wave formation in fluids,[43] and the development of structure-preserving algorithms for fluid simulation which retain important symmetries such as conservation of energy and Kelvin's circulation theorem.[44]

3.2 A Modern Formulation of Light Transport

In this section, we will establish modern foundations for light transport theory. We will show how the central equations and quantities can be derived from Maxwell's equations, and we study its properties using exterior calculus. The impatient reader may want to refer to the summary and discussion of our results in Chapter 3.2.8 before studying the technical details.

3.2.1 The Physics of Light Transport

Can light transport be considered as an asymptotic limit of Maxwell equations? This question will concern us in the present section, and following the literature we will outline a rigorous derivation from electromagnetic theory. This will naturally lead to a Hamiltonian formulation of light transport, and using measurements, as did Lambert, we will later unveil the connection between the modern theory and classical radiometry. The setting studied in the following will be those of a spatially inhomogeneous medium with varying refractive index, partly to obtain a transport equation for such environments which was missing until now in the computer graphics literature, and partly because the geometric structure is largely obscured in the homogeneous case—as the three dimensional shape of an object is obscured in a two dimensional image.

Overview In the literature, the light transport equation is often derived using transport assumptions, and photon counts are then employed to associate physical significance to otherwise abstract concepts, for example to identify particle density with radiance.[45] Such derivations have intuitive appeal. However, rigorous arguments using photons require a treatment using quantum electrodynamics, an approach which is impractical at macroscopic length scales and unnecessary when quantum and interference effects are negligible.[46] Maxwell's

---

[42]Agrachev and Sarychev, "Solid Controllability in Fluid Dynamics".

[43]Khesin and Misiolek, "Shock Waves for the Burgers Equation and Curvatures of Diffeomorphism Groups".

[44]Pinkall, Springborn, and Weißmann, "A new doubly discrete analogue of smoke ring flow and the real time simulation of fluid flow"; Mullen et al., "Energy-Preserving Integrators for Fluid Animation"; Pavlov et al., "Structure-preserving discretization of incompressible fluids"; Weißmann and Pinkall, "Filament-based smoke with vortex shedding and variational reconnection"; de Witt, "Fluid Simulation in Bases of Laplacian Eigenfunctions"; De Witt, Lessig, and Fiume, "Fluid simulation using Laplacian eigenfunctions".

[45]See for example the classic text by Duderstadt and Martin (*Transport Theory*). Arvo ("Transfer Equations in Global Illumination") presented a derivation adapted for computer graphics.

[46]See for example the forceful argument in (Mishchenko, "Gustav Mie and the fundamental concept of electromagnetic scattering by particles: A perspective"). Common misconceptions

equations and their asymptotic limit have therefore to be employed to provide a rationale for light transport theory.

In the following, we will outline a rigorous derivation of light transport from electromagnetic theory. The physical foundations of scattering at surfaces will thereby be disregarded, since the problem is most often studied independently of transport phenomena so that a discussion of the vast literature[47] and an attempt to relate it to transport models would lead us too far afield. However, the description of scattering at surfaces using the mathematical language employed by us will be discussed in detail in Chapter 3.2.7.

The derivation of light transport theory employed by us will provide a firm theoretical basis for other approaches that have been considered before in the literature, and we will comment on these connections in detail at the end of the section with the technical details in hand.

Classical Approaches in the Literature The first derivations of light transport theory from Maxwell's equations appeared in the 1960s in plasma physics, assuming a continuous medium with spatially varying refractive index field and possibly random fluctuations in the field causing scattering,[48] and shortly afterwards the question also began to be studied in applied mathematics.[49] In the 1970s, apparently unaware of the work in plasma physics but following similar ideas, Wolf and co-workers in theoretical optics investigated the connection between radiance and Maxwell's equations using the theory of partially coherent radiation. One of their principal results was that radiance, with the usual properties attributed to it, can only exist at the short wavelength limit.[50] The radiative transfer equation and scattering received comparably

---

about photons are discussed in (Scully and Sargent, "The concept of the photon"; Kidd, Ardini, and Anton, "Evolution of the Modern Photon").

[47]See for example the comprehensive treatment by Tsang and Kong (*Scattering of Electromagnetic Waves*) or (Saillard and Sentenac, "Rigorous solutions for electromagnetic scattering from rough surfaces"; Shchegrov, Maradudin, and Méndez, "Multiple Scattering of Light from Randomly Rough Surfaces"; Elfouhaily and Guérin, "A critical survey of approximate scattering wave theories from random rough surfaces").

[48]A well known account is those by Pomraning (*The Equations of Radiation Hydrodynamics*, Chapter 5) which is based on earlier work in ("Eikonal Method in Magnetohydrodynamics"; *The Theory of Plasma Waves*; "Radiative Transfer in Dispersive Media"). Alternative approaches are for example (Stott, "A transport equation for the multiple scattering of electromagnetic waves by a turbulent plasma"; Watson, "Multiple Scattering of Electromagnetic Waves in an Underdense Plasma"; Peacher and Watson, "Doppler Shift in Frequency in the Transport of Electromagnetic Waves through an Underdense Plasma"; Watson, "Electromagnetic Wave Scattering within a Plasma in the Transport Approximation"; Besieris, "Propagation of frequency-modulated pulses in a randomly stratified plasma").

[49]Lau and Watson, "Radiation Transport along Curved Ray Paths"; Barabanenkov et al., "Application of the theory of multiple scattering of waves to the derivation of the radiation transfer equation for a statistically inhomogeneous medium"; Howe, "On the Kinetic Theory of Wave Propagation in Random Media".

[50]See (Walther, "Radiometry and Coherence"; Wolf, "Coherence and Radiometry"; Friberg, *Selected Papers on Coherence and Radiometry*) and references therein. For the theory of partially coherent radiation see for example (Mandel and Wolf, *Optical Coherence and Quantum Optics*; Wolf, *Introduction to the Theory of Coherence and Polarization of Light*), and (Wolf, "The Development of Optical Coherence Theory") for a historical account.

little attention by these researchers.[51] Complementary to the above continuum approaches, transport media consisting of a large number of randomly distributed, discrete scattering particles have also been considered. These models are typically studied by solving Maxwell's equations under idealized boundary conditions for a single or a small number of scattering particles, and the result is then extended to larger domains and to include multiple scattering. A classical treatment of the approach can be found in the treatise by Tsang and Kong,[52] and a comprehensive theory was recently developed by Mishchenko.[53]

Microlocal Analysis and Quadratic Observables   Motivated by the development of improved mathematical techniques to analyse the asymptotic behaviour of observables which are quadratic in field variables,[54] the derivation of light transport theory has recently again received considerable attention in the applied mathematics literature.[55] The work refines concepts from microlocal and semi-classical analysis where partial differential equations are studied by "lifting" them from configuration to phase space.[56] There, the "lifted" problem can be studied "microlocally" using the additional momentum variable, and often one is interested in the "semi-classical" limit when an appropriate scale parameter such as Planck's constant goes to zero, , see Fig. 3.2 for a schematic

---

Ideas similar to those by Wolf and co-workers were considered in the Russian literature by Dolin ("Beam Description of Weakly Inhomogeneous Wave Fields") and Ovchinnikov and Tartarskii ("On the problem of the relationship between coherence theory and the radiation-transfer equation").

[51]See however (Wolf, "New Theory of Radiative Energy Transfer in Free Electromagnetic Fields"; Zubairy and Wolf, "Exact Equations for Radiative Transfer of Energy and Momentum in Free Electromagnetic Fields"; Kim and Wolf, "Propagation law for Walther's first generalized radiance function and its short-wavelength limit with quasi-homogeneous sources") and scattering was considered in (Greffet and Nieto-Vesperinas, "Field theory for generalized bidirectional reflectivity: derivation of Helmholtz's reciprocity principle and Kirchhoff's law") where also derivations in the radiometry literature, for example of Kirchhoff's law, are criticized.

[52]Tsang and Kong, *Scattering of Electromagnetic Waves*.

[53]See ("Maxwell's equations, radiative transfer, and coherent backscattering: A general perspective"; *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering*; "Radiative Transfer: A New Look of the Old Theory"; "Gustav Mie and the fundamental concept of electromagnetic scattering by particles: A perspective"; "Electromagnetic scattering by a morphologically complex object: Fundamental concepts and common misconceptions") and references therein.

[54]Observables are here to be understood as observable for a quasi-stationary system, cf. Remark 2.148.

[55]See for example the recent article by Ryzhik, Papanicolaou, and Keller ("Transport equations for elastic and other waves in random media") and references therein. This work was further considered by Bal ("Transport through Diffusive and Nondiffusive Regions, Embedded Objects, and Clear Layers"; "Radiative transfer equations with varying refractive index: a mathematical perspective"; "Transport Approximations in Partially Diffusive Media").

[56]For introductions to microlocal and semi-classical analysis see for example (Arnold, *Lectures on Partial Differential Equations*; Martinez, *An Introduction to Semiclassical and Microlocal Analysis*; Wunsch, *Microlocal Analysis and Evolution Equations*; Guillemin and Sternberg, *Semi-Classical Analysis*; Evans and Zworski, *Semiclassical Analysis*). Early work on the subject in the Russian literature was also done by Maslov (*Théorie des Perturbations et Méthodes Asymptotiques*), see also (Arnold, *Mathematical Methods of Classical Mechanics*, Appendix 11).

Microlocal Analysis   Electromagnetism

space of (pseudo-) differential operators

configuration space

$$\frac{\partial}{\partial t}\begin{pmatrix}\vec{E}\\ \vec{H}\end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{\epsilon}\nabla\times \\ \frac{1}{\mu}\nabla\times & 0 \end{pmatrix}\begin{pmatrix}\vec{E}\\ \vec{H}\end{pmatrix}$$

fully equivalent

Wigner transform

space of symbols

phase space cotangent bundle T^*Q

$$-\frac{\partial W^\varepsilon}{\partial t} = \frac{1}{\varepsilon}\tilde{p}^\varepsilon W^\varepsilon + \frac{1}{i}\{p^\varepsilon, W^\varepsilon\} + O(\varepsilon)$$

$\varepsilon \to 0$

principal symbol

cosphere bundle S^*Q

$$\frac{\partial W_a^0}{\partial t} + \{\tau_a, W_a^0\} = \left[W_a^0, F_a\right]$$

zeros of principal symbol

Hamiltonian

$$\dot{\ell} = -\{\ell, H\}$$
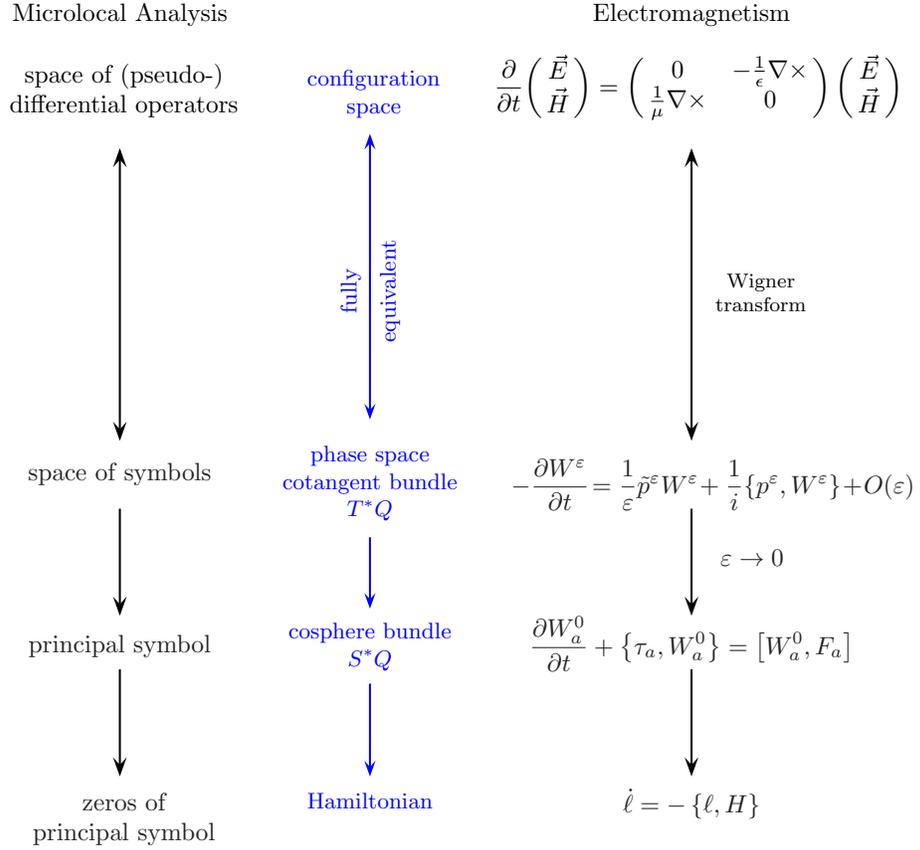
Figure 3.2: Derivation of radiative transfer from electromagnetic theory using the Wigner transform, and comparison to the general theory of microlocal analysis.

depiction.[57] The cotangent bundle T^*Q arises in these theories as the space on which the symbols $\sigma(q,p) \in \mathcal{F}(T^*Q)$ of partial differential operators are defined,[58] and the fiber variable $p \in T_q^*Q$ can intuitively be understood as a generalization of the eigenvalues for finite operators or of the spectrum for Fourier multipliers,[59] with the fibered structure of the space being necessary since the operators are spatially inhomogeneous. The polynomial order n of the symbol $\sigma(q,p)$ is the order of the original partial differential equation, and for example for Maxwell's equations which are first order the symbol has terms constant and linear in the momentum. The highest order term of the

---

[57]That classical mechanics is the limit of quantum mechanics as $\hbar \to 0$ is sometimes referred to as Bohr correspondence principle.

[58]More precisely, the symbols are elements in the coordinate ring of T^*Q, that is in the quotient $\mathcal{P}(T^*Q)/I_{\mathcal{P}}^0(T^*Q)$ where $\mathcal{P}(T^*Q)$ is the polynomial ring over T^*Q and $I_{\mathcal{P}}^0(T^*Q)$ is its ideal formed by all polynomials vanishing on T^*Q.

[59]Fourier multipliers are referred to as 'filters' in the engineering literature.

symbol $\sigma(q, p)$ is known as the principal symbol $\sigma_n(q, p)$, and it is of particular importance because of its covariance, which does not hold for the other terms of $\sigma(q, p)$. Additionally, the principal symbol $\sigma_n(q, p)$ is a homogeneous function on the cotangent bundle T^*Q,[60] which, analogous to the construction of the contact structure in Chapter 2.3.4.5, allows to identify it with a function on the cosphere bundle S^*Q.[61] For first order systems, this enables to reduce the dynamics from T^*Q to S^*Q. The zeros of the principal symbol define the characteristics of the partial differential equation, a set of ordinary differential equations that determines the time evolution for point solutions on the characteristic hyper-surface.

When a commutator bracket describes the time evolution of the original system, then the lift to the cotangent bundle allows to identify the commutator with the canonical Poisson bracket on the space of symbols T^*Q in the semi-classical regime when the scale parameter goes to zero, and time evolution is then a Hamiltonian flow on the cotangent bundle. In fact, much of microlocal and semi-classical analysis was developed for the quantization problem to bridge the divide between quantum mechanics, where dynamics are described by commutators, and classical mechanics, with the Poisson bracket determining time evolution, see Fig. 3.3 and compare also again to Fig. 3.2

The concepts and ideas from microlocal and semi-classical analysis have recently been employed to study the asymptotic behaviour of observables quadratic in oscillatory field variables $\phi^\varepsilon \in \mathcal{T}_s^r(Q)$ for a small parameter ε, such as the Planck constant, when the time evolution of the fields ϕ^ε is described by partial differential equations, cf. Chapter 2.3.4.4. An example for such a field is the energy or position density

$$n^\varepsilon(q, t) = A\left(\phi^\varepsilon(q, t), \phi^\varepsilon(q, t)\right) = \langle A\phi^\varepsilon(q, t), \phi^\varepsilon(q, t) \rangle \tag{3.1}$$

where A is a suitable bilinear form, and for simplicity we will restrict ourselves for the moment to sections $\phi^\varepsilon \in \mathcal{F}(Q)$, see Appendix A for the case of arbitrary tensor bundles. Determining the observables $n^0(q, t)$ at the limit $\varepsilon \to 0$ is nontrivial because the limit data $n^0(q, t_0)$ at the initial time t_0 is insufficient to compute $n^0(q, t)$ at later times. Hence, the propagation of the fields ϕ^ε has to be considered. However, rather than studying the propagation directly, it is convenient to employ a representation that is more suitable to determine asymptotic limits of quadratic observables. Such a representation is provided by the Wigner transform $W : \mathcal{F}(Q) \to \mathcal{F}(T^*Q)$, which, in the spirit of microlocal analysis, "lifts" a problem from configuration to phase space. For field variables ϕ^ε it is given by

$$w_\phi^\varepsilon(q, p, t) = W(\phi^\varepsilon(q, t)), \tag{3.2}$$

---

[60]Strictly speaking, since the symbols are elements in a quotient of the polynomial ring over T^*Q, the principal symbol is a homogeneous polynomial.

[61]The space of homogeneous polynomials is isomorphic to the space of polynomials on the sphere (Freeden, Gervens, and Schreiner, *Constructive Approximation on the Sphere (With Applications to Geomathematics)*) and, by the Weierstrass approximation theorem, the space of polynomials is dense in the space of continuous functions, cf. also (Atiyah, Patodi, and Singer, "Spectral asymmetry and Riemannian geometry. III", p. 79).

Quantum Mechanics   Electromagnetism

density operator, $\Phi, [\,,]$ "configuration space" $\mathcal{U}(q)$ electromagnectic
von Neumann   field
equation

Wigner   fully   Weyl
transform   equivalent   quantization

phase space
quantum mechanics $W, \{\,,\}_{\mathrm{MB}}$ phase space W optical coherence
(Moyal bracket)   Wigner distribution   theory
                                                 (with scale $\hbar/\varepsilon$)

Gronewold- $\hbar \to 0$ $\varepsilon \to 0$
van Hove
theorem

"macroscopic"

classical   light
mechanics   transport

$\ell(q,p), \{\,,\}$

symplectic diffeomorphisms

Figure 3.3: Comparison of the derivation of "inverse quantization" and light transport.
"Inverse quantization" lifts from the better known Hilbert space formulation of quantum
mechanics to the phase space formulation, and it leads to classical mechanics when
the limit $\hbar \to 0$ of the Planck constant is considered.

and w_ϕ^ε is known as Wigner quasi probability distribution. The usefulness of
the Wigner transform to study the limiting behaviour with respect to a scale
parameter has been known for quite some time,[62] and it is well suited to study

---

[62]Wigner introduced the transform named after him to study the connection between
quantum and classical mechanics (*Gruppentheorie und ihre Anwendungen auf die Quanten-
mechanik der Atomspektren*; "On the Quantum Correction For Thermodynamic Equilibrium")
although one should bear in mind that it is by no means unique and a family of closely related
mappings from configuration to phase space exists (Evans and Zworski, *Semiclassical Analysis*,
Sec. 4.1). The Wigner transform was first considered in detail in the 1940s when quantum
mechanics on phase space was developed (Moyal, "Quantum mechanics as a statistical theory";
Groenewold, "On the principles of elementary quantum mechanics") which provides a fully
independent formulation of quantum mechanics, next to the better known Hilbert space and
path integral formulations, see (Zachos, Fairlie, and Curtright, *Quantum Mechanics in Phase
Space: An Overview with Selected Papers*). The limit $\hbar \to 0$ of the phase space formulation of

quadratic observables since these can be recovered directly from the Wigner quasi probability distribution $w_\phi^\varepsilon(q, p, t)$ by the fiber integral

$$n^\varepsilon(q, t) = \int_{T_q^* Q} w^\varepsilon(q, p, t) \, dp. \tag{3.3}$$

The above equality holds under suitable technical assumptions also asymptotically as $\varepsilon \to 0$, and establishing this result under realistic regularity assumptions on the field variable ϕ^ε was the principal accomplishment of recent work where the Wigner transform was employed. The limit Wigner distribution $w^0 = \lim_{\varepsilon \to 0} w^\varepsilon$ is known under various names, such as H-measure, microlocal defect measure, semi-classical measure, and Wigner measure, cf. Remark 3.5 for a discussion of the relationship between these concepts, and the nomenclature is justified since it is guaranteed to be positive, which does not hold for the Wigner quasi probability distribution which can attain negative values, cf. Appendix 3.2.8. For higher order quadratic forms of the field variables ϕ^ε, such as the flux density

$$\vec{N}^\varepsilon(q, t) = \langle D^i \phi^\varepsilon(q, t), \phi^\varepsilon(q, t) \rangle \tag{3.4}$$

where D^i is a differential operator for the ith coordinate, the Wigner distribution provides the analogous benefits, and for example the asymptotic flux density $\vec{N}^0$ is the first moment

$$\vec{N}^0(q, t) = \int_{T_q^* Q} \frac{p}{\|p\|} \, w^0(q, p, t) \, dp \tag{3.5}$$

of the Wigner distribution. Although the Wigner transform provides a convenient representation of the fields ϕ^ε on phase space, its usefulness for initial value problems relies on a well defined lift of the dynamics to phase space at the limit $\varepsilon \to 0$. This problem has been studied in detail for the Schrödinger equation

$$0 = i\hbar \frac{\partial \phi}{\partial t} + \frac{\hbar}{2} \Delta \phi - V(q)\phi \tag{3.6a}$$

which using the Hamiltonian

$$\hat{H} = \hbar/2 \, \Delta - V(q) \tag{3.6b}$$

can be written as

$$0 = i\hbar \frac{\partial \phi}{\partial t} + \hat{H}\phi \tag{3.6c}$$

and the equation often serves as model for these problems because of its importance for quantization.[63] Rescaling the units in Eq. 3.6 to obtain $\hbar$-dependent wave functions $\phi^\hbar$ and using the tools from microlocal analysis and

---

quantum mechanics leads to classical mechanics, see again Fig. 3.3.

[63]For situations where the lift is not well defined, which are largely related to spatial inhomogeneities that lead to a qualitative change or to nonlinearities, see (Carles et al., "On the time evolution of Wigner measures for Schrödinger equations"); cf. also the discussion in (Emmrich and Weinstein, "Geometry of the transport equation in multicomponent WKB approximations").

324

the calculus of symbols to lift the Hamiltonian $\hat{H}$ from configuration to phase space, the Schrödinger equation becomes on T^*Q at the semi-classical limit

$$\frac{\partial w^0}{\partial t} + p \cdot \nabla_q w^0 - \nabla_q V(q) \cdot \nabla_p w^0 = 0 \tag{3.7}$$

which is easily recognized as the Vlasov equation

$$\frac{\partial w^0}{\partial t} + \left\{ w^0, H \right\} = 0 \tag{3.8}$$

for the Hamiltonian

$$H(q, p) = \|p\|^2 + V(q) \tag{3.9}$$

given by the symbol of the quantum Hamiltonian $\hat{H}$ in Eq. 3.6b, cf. Chapter 2.3.4.3.

Using the above methodology, we will in the following outline how the transport of the electromagnetic energy density, which is quadratic in the electric and magnetic field variables, leads at the short wavelength limit to light transport theory. At the end of the derivation, with the technical details in hand, we will also comment again on the connection to other derivations of the light transport equation.

From Electromagnetic Theory to Radiative Transfer[64]   Let the configuration space $Q \subset \mathbb{R}^3$ be a smooth, compact manifold with the usual Euclidean metric, or $Q = \mathbb{R}^3$ with suitable decay conditions on the tensor fields at infinity, cf. Remark 2.120, and consider a time-harmonic electromagnetic field on Q with support in the visible frequency range $\nu \in [4 \times 10^{-14}, 7.5 \times 10^{-14}]$ Hertz.[65] Using geometic units, Maxwell's equations describing the time evolution of the electric and magnetic vector fields $\vec{E}(q, t) \in \mathfrak{X}(Q)$ and $\vec{H}(q, t) \in \mathfrak{X}(Q)$, are

$$\epsilon \frac{\partial \vec{E}}{\partial t} = \nabla \times \vec{H} \tag{3.10a}$$

$$-\mu \frac{\partial \vec{H}}{\partial t} = \nabla \times \vec{E} \tag{3.10b}$$

and we assume that the constitutive fields $\epsilon(q) \in \mathcal{F}(Q)$ and $\mu(q) \in \mathcal{F}(Q)$ which characterize the transport medium are isotropic and time-invariant. The

---

[64]Our derivation follows Gérard, Markowich, Mauser, and Poupaud ("Homogenization limits and Wigner transforms"). A less precise but somewhat more intuitive derivation was presented by Ryzhik, Papanicolau, and Keller ("Transport equations for elastic and other waves in random media"), with some additional details available in (Papanicolaou and Ryzhik, "Waves and Transport"). An outline of the ansatz can also be found in (Bal, "Radiative transfer equations with varying refractive index: a mathematical perspective"), which was the starting point for our investigations on the subject.

[65]See for example the classic text by Born and Wolf (*Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*) for the necessary background on electromagnetic theory.

dynamics in Eq. 3.10 are subject to the divergence freeness constraints[66]

$$\operatorname{div}\left(\epsilon\,\vec{E}\right) = 0 \qquad \operatorname{div}\left(\mu\,\vec{H}\right) = 0. \tag{3.11}$$

Eq. 3.10 can be written more concisely as

$$\frac{\partial U}{\partial t} + P(q)\,U = 0 \tag{3.12}$$

where $U = (\vec{E}, \vec{H})^T$ represents the electromagnetic vector fields. The first order, partial differential "Maxwell operator" $P(q)$ is defined as

$$P(q) = DA^{-1} = \begin{pmatrix} 0 & -\frac{1}{\epsilon}\nabla\times \\ \frac{1}{\mu}\nabla\times & 0 \end{pmatrix} \tag{3.13}$$

where D is the differential operator expressing the curl of the fields and $A(q) \in \mathcal{T}_2^0(Q)$ is a quadratic bilinear form whose components a_{ij} are given by

$$a_{ij} = \epsilon(q)\delta_{ij} \qquad 1 \leq i \leq 3 \tag{3.14a}$$

$$a_{ij} = \mu(q)\delta_{ij} \qquad 4 \leq i \leq 6 \tag{3.14b}$$

and which can be seen as the natural metric for electromagnetic theory.[67]

Remark 3.1. The form of Maxwell's equations in Eq. 3.12 and the Maxwell operator $P(q)$ in Eq. 3.13 arise naturally when electromagnetic theory is consider in space-time using differential forms.[68] Maxwell's equations for a source free electromagnetic field then take the form

$$\operatorname{d}(\star F) = 0 \tag{3.15}$$

where the Hodge star operator is equivalent to the bilinear form A defined in Eq. 3.14[69] and the Faraday 2-form $F \in \Omega^2(Q \times \mathbb{R})$ is related to the differential forms $E \in \Omega^1(Q)$ and $B \in \Omega^2(Q)$ associated with the electric and magnetic fields by

$$F = E \wedge dt + B \in \Omega^2(\mathbb{R}^3 \times T). \tag{3.16}$$

---

[66]As we discussed in Example 2.139, the divergence freeness of the electric field $\vec{E}$ arises from the gauge symmetry of Maxwell's equations in the vector potential. However, we will not be concerned with these aspects in this section.

[67]This was pointed out in (Ryzhik, Papanicolaou, and Keller, "Transport equations for elastic and other waves in random media") and is also closely related to the metric defined in (Bossavit and Kettunen, "Yee-Like Schemes on Staggered Cellular Grids: a Synthesis Between FIT and FEM Approaches").

[68]See for example (Stern et al., "Geometric Computational Electrodynamics with Variational Integrators and Discrete Differential Forms", Sec. 2) for an introduction.

[69]Bossavit and Kettunen, "Yee-Like Schemes on Staggered Cellular Grids: a Synthesis Between FIT and FEM Approaches"; Stern et al., "Geometric Computational Electrodynamics with Variational Integrators and Discrete Differential Forms".

Hence, the components of the Faraday 2-forms F are given by $U = (\vec{E}, \vec{H})^T$ and the spatial part of the space-time operator $\mathrm{d}\star$ in Eq. 3.15 is nothing but the Maxwell operator $P(q)$. Note that the form of Maxwell's equations in Eq. 3.15 is the Euler-Lagrange equation for the Lagrangian $L = -1/2\,(\star F \wedge F)$ and Poynting's theorem, describing the conservation of the electromagnetic energy density, is easily obtained from the Lagrangian as

$$0 = -2\,\mathrm{d}L = \mathrm{d}(\star F \wedge F) = \mathrm{d}(\star F) \wedge F \tag{3.17}$$

where we used the Leibniz rule for differential forms and that F is exact.[70]

In many situations, the constitutive fields $\epsilon(q)$ and $\mu(q)$ fluctuate over a characteristic distance d_m much larger than the wavelength of the radiation λ. With the ratio of the two length scales being

$$\varepsilon = \frac{\lambda}{d_m} \tag{3.18}$$

one can introduce scaled variables $q \to \varepsilon q$ and $t \to \varepsilon t$. Eq. 3.12 then becomes

$$\frac{\partial U^\varepsilon}{\partial t} + P^\varepsilon(q)\,U^\varepsilon = 0 \tag{3.19}$$

where $P^\varepsilon(q)$ and U^ε are the scaled Maxwell operator and fields, respectively.

We are interested in the short wavelength limit of Eq. 3.19 when $\varepsilon \to 0$ and for this regime we seek the time evolution of quadratic observables such as the energy density

$$\mathcal{E}(q,t) = \frac{\epsilon}{2}\|\vec{E}\|^2 + \frac{\mu}{2}\|\vec{H}\|^2 \tag{3.20a}$$

$$= \frac{1}{2}A(U^\varepsilon, U^\varepsilon) \tag{3.20b}$$

$$= \frac{1}{2}\langle U^\varepsilon, U^\varepsilon \rangle_A \in \mathrm{Den}(Q) \tag{3.20c}$$

and the flux density

$$\vec{S}_i(q,t) = \vec{E} \times \vec{H} \tag{3.21a}$$

$$= \frac{1}{2}A(D^i U^\varepsilon, U^\varepsilon) \tag{3.21b}$$

$$= \frac{1}{2}\langle D^i U^\varepsilon, U^\varepsilon \rangle_A \in \mathfrak{X}(Q) \tag{3.21c}$$

where A is again the bilinear form defined in Eq. 3.14, cf. Eq. 3.1 and Eq. 3.4. As discussed before, $\mathcal{E}(q,t)$ and $\vec{S}(q,t)$ cannot be obtained directly from $\mathcal{E}(q,0)$ and $\vec{S}(q,0)$ at the initial time $t = 0$. Hence, we will employ the Wigner transform to lift the electromagnetic field from configuration to phase space, and study the

---

[70]Cf. (Kiehn and Pierce, "Intrinsic Transport Theorem").

short wavelength limit and the time evolution of the quadratic observables in Eq. 3.20 and Eq. 3.21 there. In the following, we will have to assume a working knowledge of the Wigner transform, and the reader unfamiliar with the concept is referred to Appendix A.

For the ε-dependent electromagnetic vector field U^ε in Eq. 3.19, the Wigner transform yields a 6×6 Hermitian "matrix" density $W^\varepsilon(q, p) = W_U^\varepsilon(q, p)$ on phase space whose elements are[71]

$$W_{ij}^\varepsilon(q,p) = \frac{1}{(2\pi)^3} \int_Q e^{ip \cdot r} \, U_i^\varepsilon \Big(q - \frac{\varepsilon}{2}r\Big) \, U_j^\varepsilon \Big(q + \frac{\varepsilon}{2}r\Big) \, dr. \qquad (3.22)$$

The physical significance of the momentum variable in Eq. 3.22 can be illuminated with the definition of the Wigner transform using the Fourier transform. In the literature, the Fourier transform of the electromagnetic fields U^ε is known as angular spectrum representation[72] and it can be interpreted as the superposition of an infinite number of plane waves of the form

$$U(r, t) = A e^{i(\vec{k} \cdot r - \omega t)} \qquad (3.23)$$

where $\vec{k}$ is the wave vector, ωt represents the time harmonic part of the field, and $r \in Q$.[73] A comparison of the plane wave equation in Eq. 3.23 with the Wigner transform in Eq. 3.22 shows that the momentum p is equivalent to the wave vector $\vec{k}$. In the following, we will continue to write 'p' for the momentum to emphasize the connection to Hamiltonian mechanics that will arise in the following, but it should be kept in mind that its physical significance is given by the expression for the wave vector, that is

$$p = \vec{k}^\flat = \Big(\frac{n(q)}{c}\omega\bar{s}\Big)^\flat = \Big(\frac{n(q)}{c}(2\pi\nu)\,\bar{s}\Big)^\flat \qquad (3.24)$$

where $\bar{s}$ is a unit vector.

To lift Maxwell's equations in the operator form in Eq. 3.19 to phase space we consider the Wigner distribution as a function of $\vec{U}^\varepsilon$ and differentiate with respect to time. By the Leibniz rule this yields

$$\frac{\partial W^\varepsilon}{\partial t} = \frac{\partial}{\partial t} W^\varepsilon(U^\varepsilon, U^\varepsilon) = W^\varepsilon \Big(\frac{\partial U^\varepsilon}{\partial t}, U^\varepsilon\Big) + W^\varepsilon \Big(U^\varepsilon, \frac{\partial U^\varepsilon}{\partial t}\Big) \qquad (3.25a)$$

and using Eq. 3.19 we obtain

$$\frac{\partial W^\varepsilon}{\partial t} = \frac{\partial}{\partial t} W^\varepsilon(U^\varepsilon, U^\varepsilon) = W^\varepsilon(-P^\varepsilon U^\varepsilon, U^\varepsilon) + W^\varepsilon(U^\varepsilon, -P^\varepsilon U^\varepsilon). \qquad (3.25b)$$

---

[71]A closely related matrix was apparently already studied by Wiener ("Generalized Harmonic Analysis", Chapter 9) in his famous paper on generalized harmonic analysis, cf. (Wolf, "The influence of Young's interference experiment on the development of statistical optics", Sec. 4).

[72]See for example (Mandel and Wolf, *Optical Coherence and Quantum Optics*, Chapter 3.2) and an introduction can for example be found in (Cuypers et al., "Validity of Wigner Distribution Function for ray-based imaging", Sec. 3).

[73]Born and Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, p. 14.

We can simplify the equation using[74]

$$W^\varepsilon(S\Phi^\varepsilon, \Phi^\varepsilon) = \sigma\, W^\varepsilon_\Phi + \frac{\varepsilon}{2i}\left\{\sigma, W^\varepsilon_\Phi\right\} + O(\varepsilon^2) \tag{3.26a}$$

$$W^\varepsilon(\Phi^\varepsilon, S\Phi^\varepsilon) = W^\varepsilon_\Phi\, \sigma^* + \frac{\varepsilon}{2i}\left\{W^\varepsilon_\Phi, \sigma^*\right\} + O(\varepsilon^2) \tag{3.26b}$$

where S is a differential operator with symbol σ and adjoint symbol σ^*, Φ^ε is a vector valued field with Wigner distribution W^ε_Φ, and $\{\,,\}$ is the canonical Poisson bracket on the cotangent bundle which is considered component-wise. With Eq. 3.26, we obtain for Eq. 3.25b thus

$$-\frac{\partial W^\varepsilon}{\partial t} = \frac{1}{\varepsilon}\left(p^\varepsilon\, W^\varepsilon - W^\varepsilon p^\varepsilon\right) + \frac{1}{2i}\left(\{p^\varepsilon, W^\varepsilon\} - \{W^\varepsilon, p^\varepsilon\}\right) + O(\varepsilon) \tag{3.27}$$

where p^ε is the symbol of the Maxwell operator P^ε and we also used that it is an essentially self-adjoint operator,[75] which follows from P^ε being real. Expanding the Maxwell symbol $p^\varepsilon = p^0 + \varepsilon p^1 + O(\varepsilon^2)$ in orders of ε and with the anti-symmetry of the Poisson bracket we obtain

$$-\frac{\partial W^\varepsilon}{\partial t} = \frac{1}{\varepsilon}\left(p^0 W^\varepsilon - W^\varepsilon p^0\right) + \left(p^1 W^\varepsilon + W^\varepsilon p^1\right) + \frac{1}{i}\left\{p^\varepsilon, W^\varepsilon\right\} + O(\varepsilon). \tag{3.28}$$

We are interested in the asymptotic solutions as $\varepsilon \to 0$. However, the first term in the equation will "blow up" at the limit. Hence, any solution has to satisfy

$$p^0\, W^0 - W^0\, p^0 = 0 \tag{3.29}$$

and the zero order term p^0 of the symbol and the limit Wigner distribution W^0 have to commute. The Wigner distribution W^0 represents an electromagnetic field that is arbitrary. Hence, the limit Maxwell symbol p^0 has to lie in the ideal of the group of square matrices W^0 so that it commutes with any Wigner transformed field W_0, that is, p^0 has to be a scalar matrix. But p^0 is scalar only on each of its eigenspaces where it takes the form $\hat{p}^a_{ij} = \tau_a \delta_{ij}$, where τ_a is the eigenvalue and the size of $\hat{p}^a$ is defined by the multiplicity m_a of τ_a. Eq. 3.28 is therefore only satisfied when W^0 is restricted to the eigenspaces of the zero order Maxwell symbol p^0. Eq. 3.28 then leads to independent transport equation for each eigenspace, and at the limit $\varepsilon \to 0$ these are to first order given by[76]

$$\frac{\partial W^0_a}{\partial t} + \left\{\tau_a, W^0_a\right\} - \left[W^0_a, F_a\right] = 0 \tag{3.30}$$

[74] A proof can be found in the appendix of (Gérard et al., "Homogenization limits and Wigner transforms") and the result is a consequence of the general calculus of pseudo-differential operators, see for example (Hörmander, *The Analysis of Linear Partial Differential Operators*).

[75] See for example (Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*, Chapter 8.4) for a discussion of essentially self-adjoint operators.

[76] The details of the derivation can be found in (Gérard et al., "Homogenization limits and Wigner transforms", Sec. 6). The ansatz is to multiply Eq. 3.28 from the left and right by Π_a and employ $p^0 = i\sum_a \lambda_a \Pi_a$ in the terms with the Poisson bracket where the Maxwell symbol has to be interpreted as a function. This enables to exploit linearity so that only the eigenvalue is left in the bracket.

where $W_a^0 = \Pi_a W^0 \Pi_a$ is the $m_a \times m_a$ "matrix" obtained by the projection Π_a onto the eigenspace of τ_a, and $F_a = [\Pi_a, \{\tau_a, \Pi_a\}] + \Pi_a\, p^1 \Pi_a$ with $[\,,]$ the usual commutator for operators, and the second and third term in Eq. 3.30 both arise from $\frac{1}{i}\{p^\varepsilon, W^\varepsilon\}$ in Eq. 3.28.

At the asymptotic limit, the eigenvalues of the Maxwell symbol p^0 have all multiplicity two and are given by

$$\tau_0(q, p) = 0 \tag{3.31a}$$

$$\tau_1(q, p) = \frac{c}{n(q)}\|p\| \tag{3.31b}$$

$$-\tau_2(q, p) = \frac{c}{n(q)}\|p\|. \tag{3.31c}$$

The eigenvalues relevant for energy transport are τ_1 and τ_2, which intuitively correspond to forward and backward propagation in time, and by their symmetry only one of them has to be considered. Projecting the limit Wigner distribution W^0 onto the eigenspace of τ_1 and identifying the eigenvectors spanning the space with the directions associated with the Stokes parameters I, Q, U, V for polarized light,[77] one obtains

$$\alpha(q, p) = \Pi_1 W^0(q, p)\Pi_1 = \frac{1}{2}\left[\begin{array}{cc} I + Q & U + iV \\ U - iV & I - Q \end{array}\right] dq \wedge dp \tag{3.32}$$

and $\alpha(q, p)$ is known in the literature as phase space coherence density or coherence matrix. By Eq. 3.30, the transport equation for the phase space coherence density $\alpha(q, p)$ is given by

$$\frac{\partial \alpha}{\partial t} + \{\tau_1, \alpha\} + F_1\alpha - \alpha F_1 = 0 \tag{3.33}$$

where F_1 can be interpreted as describing the rotation of the polarization during transport, and the vanishing right hand side signifies the neglect of scattering effects.

When only unpolarized radiation is considered one has $Q = U = V = 0$ and α becomes a scalar matrix. Hence the commutator $F_1\alpha - \alpha F_1$ in Eq. 3.33 vanishes, and using components it is also easy to show that in this case taking the trace commutes with the remaining transport equation. For unpolarized radiation the transport equation in Eq. 3.33 hence becomes

$$\frac{\partial \ell}{\partial t} + \{\tau_1, \ell\} = 0 \tag{3.34}$$

where the **phase space light energy density** ℓ is defined by

$$\ell = \mathrm{tr}\,(\alpha(q, p)) = \mathcal{L}(q, p)\, dq \wedge dp \in \mathrm{Den}(T^*Q) \tag{3.35}$$

---

[77] The four Stokes parameters I, Q, U, V were introduced by Stokes in 1852 to describe the polarization of radiation. Intuitively, I represents the intensity, Q and U linear polarization for two coordinate systems that are 45 degree rotated with respect to each other, and V circular polarization. Q, U, V are all defined in a plane orthogonal to the direction of propagation.

and it is positive since the non-vanishing Stokes parameter I is positive. By the general theory of the Wigner distribution, cf. Eq. 3.191 in Appendix A, the phase space light energy density is related to the electromagnetic energy density $\mathcal{E}(q)$ by

$$\mathcal{E}(q) = \int_{T_q^*Q} \ell \, dp \tag{3.36}$$

and there is no summation over the eigenvalues since τ_1 and τ_2 yield symmetric solutions.[78] Analogously, for the flux density of the electromagnetic field, which on configuration space is given by the Poynting vector $\vec{S} = \vec{E} \times \vec{H}$, we obtain

$$\vec{S}(q) = \int_{T_q^*Q} \frac{p}{\|p\|} f \, dp. \tag{3.37}$$

Eq. 3.34 shows that the dynamics of the light energy density ℓ are described by a Hamiltonian system for the Hamiltonian function

$$H(q,p) = \tau_1(q,p) = \frac{c}{n(q)}\|p\| = \frac{c}{n(q)}\|\vec{k}\| = 2\pi\nu \tag{3.38}$$

and it is homogeneous of degree one in the momentum, as one would expect from the general theory of microlocal analysis for a first order partial differential equation such as Maxwell's equations. The time evolution of the phase space light energy density is described by the **light transport equation**

$$\frac{\partial \ell}{\partial t} = -\{\ell, H\} \tag{3.39}$$

which is in the form of the Vlasov equation that plays a central role in various areas of physics.[79] The time evolution of phase space points is given by the Hamiltonian vector field $X_H = (\dot{q}, \dot{p})$ for the Hamiltonian in Eq. 3.38. By Hamilton's equations in Theorem 2.29, it is given by

$$\frac{\partial q}{\partial t} = \frac{\partial H}{\partial p} = \frac{c}{n(q)} \frac{p}{\|p\|} \tag{3.40a}$$

$$-\frac{\partial p}{\partial t} = \frac{\partial H}{\partial q} = \nabla_q \frac{c}{n(q)}\|p\| = \nabla_q\left(\frac{c}{n(q)}\right)\|p\| = c\,\|p\|\frac{\nabla_q n(q)}{n^2(q)}. \tag{3.40b}$$

Eq. 3.40 can be understood as the particle or wave packet[80] analogue for the time evolution equation for the phase space density in Eq. 3.39. However, the

---

[78]More precisely, the sum cancels with the factor of one half in Eq. 3.20.

[79]See (Henon, "Vlasov Equation") for some historical remarks on the Vlasov equation. Physical contexts where the equation arises are discussed in (Holm, Putkaradze, and Tronci, "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions").

[80]A wave packet is the superposition of elementary waves that is localized in space and time by constructive and destructive interference, see (Harris, "Radiative Transfer in Dispersive Media") for the special case of electromagnetic radiation and (Littlejohn, "The semiclassical evolution of wave packets") for a more detailed and modern discussion of wave packets. Confusingly, Pomraning calls wave packets also photons (*The Equations of Radiation Hydrodynamics*, p. 145). The mathematical description of wave packets is again closely related to the Wigner distribution.

two formulations are not equivalent in that it is in general not possible to deduce the continuous light energy density ℓ_t from a set of discrete particles transported by the Hamiltonian vector field in Eq. 3.40b. The question under which assumptions this equivalence does hold is a principal motivation for our work in the second part of the present thesis in Chapter 4.

In the following, we will consider the central objects of light transport theory, such as the light transport equation and the phase space light energy density, in more detail and we will further investigate the Hamiltonian structure of the theory. Additionally, questions such as the measurement of light energy density and scattering will be investigated. However, before we continue some remarks are in order.

Remark 3.2. The Hamiltonian vector field X_H in Eq. 3.40 will usually only be defined locally so that there is a local flow η_t in the sense of Def. 2.91 over the domain of X_H, for example between the boundaries of an enclosure or in between surfaces. Nonetheless, almost all of the results established in Chapter 2.3 for complete flows are available with suitable modifications, for example η_t provides a local diffeomorphism and the group property $\eta_{t+s} = \eta_t \circ \eta_s$ is locally satisfied by Proposition 2.44. The rich structure that arises when X_H is defined globally will be studied in Chapter 3.3.

Remark 3.3. We discussed in the foregoing that photon counts are not a suitable physical foundation for light transport theory and that electromagnetic theory has to be employed as we did in the preceding discussion. Nonetheless, the use of a particle model for light is intuitive—in the same way as one talks about the transport of molecules in ideal fluid dynamics[81]—and Hamilton's equations in Eq. 3.40b allow to describe this idea mathematically.[82] Additionally, a derivation of the radiative transfer equation based on continuum assumptions can be illuminating to understand the principles and the effects modelled.[83]

Remark 3.4. The classical approach to study the short wavelength limit of Maxwell's equations is the eikonal or Wentzel-Kramer-Brilliouin (WKB) approximation whose use for electromagnetic theory goes back to the early 20th century when it was employed to show that geometric optics arises from Maxwell's equations at the short wavelength limit.[84] In contrast to the eikonal ansatz, the Wigner transform enables to study the short wavelength limit for

---

[81]See for example the excellent book by (Chorin and Marsden, *A Mathematical Introduction to Fluid Mechanics*).

[82]It should be mentioned at this point that the motivation for the method of characteristics in microlocal analysis lies in obtaining a particle description on phase space that is easier to study and integrate in time than the original partial differential equation.

[83]The insight that can be gained by a treatment using transport theory is beautifully illustrated in the book by Chorin and Marsden (*A Mathematical Introduction to Fluid Mechanics*) for fluid dynamics.

[84]Wentzel, Kramer, and Brilliouin developed the method that bears now their name to obtain approximations to the Schrödinger equation, although the ansatz was known at least since the times of Green and Liouville. For a derivation of geometric optics from Maxwell's equations using an eikonal approach see for example (Born and Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Chapter III).

observables quadratic in the field variables, to extend solutions beyond caustics by prolonging them in phase space,[85] to consider realistic regularity assumptions for the functions involved, and to model scattering using macroscopic models.[86] These advantages were not clear in the 1960s when the first derivations of the radiative transfer equations using the Wigner transform were proposed in plasma physics.[87] However, it appears that the transform was at the time already well known from applications in quantum mechanics, and hence a "natural" approach to obtain a description in phase space with an angular dependence. Wolf and co-workers in theoretical optics also used the Wigner transform,[88] arguably since it also plays a central role in optical coherence theory. However, they have not related their results to other contexts where the transform has been employed and hence reinvented many concepts in a restricted setting that were already known withing a more general theory.[89] For example, that radiance can only exist at the short wavelength limit is equivalent to the Wigner distribution being non-negative only in the limit when the scale parameter goes to zero, and that it is only of physical significance when integrated is readily apparent from its infinitesimal nature. The Wigner transform has recently also been employed in computer vision and computational photography and related fields.[90] The applications there are however motivated by those in theoretical optics and hence the work suffers from similar problems. For example, the Wigner transform is justified only in restricted settings but used in much more general contexts,[91] the mathematical structure underlying the transform has not been considered, and the manifold connections that the formulation opens up to other disciplines and problems have not been realized.

Remark 3.5. Since the mid 1980s, a variety of generalizations of the classical Wigner transform have been developed. H-measures or microlocal defect

---

[85]This is well known in computer graphics where caustics are simulated but the problems related to a configuration space treatment are avoided.

[86]A detailed comparison of the WKB method and the Wigner transform approach is provided in (Sparber, Markowich, and Mauser, "Wigner functions versus WKB-methods in multivalued geometrical optics"). These authors consider semi-classical analysis and symbol calculus as an extended WKB method where an infinite superposition of WKB solutions is employed, a viewpoint closely related to the angular spectrum representation found in optics.

[87]See for example (Harris, "Radiative Transfer in Dispersive Media").

[88]The Wigner transform was employed beginning with the earliest work by Walther ("Radiometry and Coherence"), see also the recent discussion of it in (Zhang and Levoy, "Wigner Distributions and How They Relate to the Light Field").

[89]This is rather surprising since Wolf was aware of the phase space formulation of quantum mechanics for which much of the general theory was developed and which is typically formulated in this language, cf. (Wolf, "Coherence and Radiometry").

[90]See for example (Zhang and Levoy, "Wigner Distributions and How They Relate to the Light Field"; Cuypers et al., "Validity of Wigner Distribution Function for ray-based imaging") and references therein.

[91]Zhang and Levoy (Zhang and Levoy, "Wigner Distributions and How They Relate to the Light Field") employ the Wigner transform for a planar setting to define radiant intensity as the magnitude of the Poynting vector. However, it is neither physically nor mathematically apparent that this generalizes to arbitrary settings. A priori it is also not possible to identify a macroscopic quantity such as radiant intensity with the microscopic magnitude of the Poynting vector.

measures were the first to appear in the literature, and these are still the most general formulations requiring the least regularity for the functions involved.[92] Shortly afterwards, semi-classical measures were introduced, and in fact it appears that only then the connection to the Wigner transform was realized.[93] Semi-classical measures forego some of the generality of H-measures and microlocal defect measures, but this enables to work with them more easily in applications.[94]

Remark 3.6. The configuration and phase space formulations of electromagnetic theory which we developed in the foregoing are fully equivalent, cf. again Fig. 3.3. Hence, an analogue of Poynting's theorem

$$\frac{\partial \mathcal{E}}{\partial t} + \nabla \cdot \vec{S} = 0 \tag{3.41}$$

that describes the transport of the energy density $\mathcal{E}(q)$ on phase space has to exist. We will in the following show that, somewhat surprisingly, this analogue is the light transport equation in Eq. 3.39 that is defined at the short wavelength limit.

With the coordinate expression for the Poisson bracket in Eq. 2.122, the integral of the light transport equation in Eq. 3.39 over an arbitrary fiber T_q^*Q is

$$0 = \int_{T_q^*Q} \frac{\partial \ell}{\partial t}\, dp + \int_{T_q^*Q} \frac{\partial H}{\partial p} \cdot \frac{\partial \ell}{\partial q}\, dp + \int_{T_q^*Q} \frac{\partial H}{\partial q} \cdot \frac{\partial \ell}{\partial p}\, dp. \tag{3.42}$$

In the first term, differentiation with respect to time and integration over momentum commute.[95] With the expressions for the partial derivatives of the Hamiltonian in Hamilton's equations in Eq. 3.40 we hence obtain

$$0 = \frac{\partial}{\partial t}\int_{T_q^*Q} \ell\, dp + \frac{c}{n(q)}\int_{T_q^*Q} \frac{p}{\|p\|} \cdot \frac{\partial \ell}{\partial q}\, dp + c\frac{\nabla n(q)}{n^2(q)} \cdot \int_{T_q^*Q} \|p\|\frac{\partial \ell}{\partial p}\, dp. \tag{3.43}$$

Using integration by parts for the last term and exploiting the decay conditions on the phase space light energy density ℓ, cf. Example 2.80, yields

$$0 = \frac{\partial}{\partial t}\int_{T_q^*Q} \ell\, dp + \frac{c}{n(q)}\int_{T_q^*Q} \frac{p}{\|p\|} \cdot \frac{\partial \ell}{\partial q}\, dp + c\frac{\nabla n(q)}{n^2(q)} \cdot \int_{T_q^*Q} \frac{p}{\|p\|}\ell\, dp. \tag{3.44}$$

[92]The theory of H-measures or microlocal defect measures was developed independently by Tartar ("H-measures and Small Amplitude Homogenization"; "H-measures, a New Approach for Studying Homogenization, Oscillations and Concentration Effects in Partial Differential Equations") and Gérard ("Microlocal defect measures"), see (Tartar, "H-measures and Applications") for motivation and also ("H-measures and Applications"; "Mathematical Tools for Studying Oscillations and Concentrations: from Young Measures to H-Measures and Their Variants"; *The General Theory of Homogenization: a personalized introduction*).

[93]Gérard, "Mesures semi-classiques et ondes de Bloch"; Lions and Paul, "Sur les mesures de Wigner".

[94]A detailed comparison between microlocal defect measures and semi-classical measures can be found in (Burq, "Mesures semi-classique et mesures de défaut").

[95]Jost, *Postmodern Analysis*, Theorem 16.10.

However, by the Leibniz rule we also have

$$\frac{\partial}{\partial q}\left(\frac{c}{n(q)}\int_{T_q^*Q}\frac{p}{\|p\|}\ell\,dp\right)$$

$$=\left(\frac{\partial}{\partial q}\frac{c}{n(q)}\right)\left(\int_{T_q^*Q}\frac{p}{\|p\|}\ell\,dp\right)+\left(\frac{c}{n(q)}\right)\left(\frac{\partial}{\partial q}\int_{T_q^*Q}\frac{p}{\|p\|}\ell\,dp\right)\quad(3.45\text{a})$$

$$=c\frac{\nabla n(q)}{n^2(q)}\cdot\int_{T_q^*Q}\frac{p}{\|p\|}\ell\,dp+\frac{c}{n(q)}\int_{T_q^*Q}\frac{p}{\|p\|}\cdot\frac{\partial\ell}{\partial q}\,dp\qquad(3.45\text{b})$$

where in the last line we used again that differentiation and integration can be interchanged.[96] With Eq. 3.45 we can write Eq. 3.44 hence as

$$0=\frac{\partial}{\partial t}\int_{T_q^*Q}\ell\,dp+\frac{\partial}{\partial q}\left(\frac{c}{n(q)}\int_{T_q^*Q}\frac{p}{\|p\|}\ell\,dp\right)\qquad(3.46)$$

But in terms of the Wigner distribution the energy density and the Poynting vector are given by[97]

$$\mathcal{E}(q)=\int_{T_q^*Q}\ell\,dp\qquad(3.47\text{a})$$

$$\vec{S}(q)=\frac{c}{n(q)}\int_{T_q^*Q}\frac{p}{\|p\|}\,\ell\,dp,\qquad(3.47\text{b})$$

cf. also Chapter 3.2.6, so that Eq. 3.46 can be written as

$$0=\frac{\partial\mathcal{E}}{\partial t}+\frac{\partial}{\partial q}\cdot\vec{S}\qquad(3.48\text{a})$$

$$=\frac{\partial\mathcal{E}}{\partial t}+\nabla\cdot\vec{S}\qquad(3.48\text{b})$$

which is again Poynting's theorem in Eq. 3.41.[98] At first sight it seems surprising that the light transport equation is equivalent to Poynting's theorem, even more so at the short wavelength limit. However, the result is anticipated in the equivalence between configuration and phase space established by the Wigner transform and Weyl map, cf. Fig. 3.3. Nonetheless, it has to be noted that the above equivalence relies on a vanishing phase space light energy density on the boundaries of the domain, an assumption which is for the compact domains of interest in applications usually not satisfied.

---

[96]The most general form of the differentiation under the integral sign theorem is needed here. Apparently, it is rarely stated in the literature, see however (Cheng, *Differentiation Under the Integral Sign with Weak Derivatives*).

[97]See (Ryzhik, Papanicolaou, and Keller, "Transport equations for elastic and other waves in random media", p. 342).

[98]The inverse result, that a continuity equation leads to the Poisson bracket, can for example be found in (Gérard et al., "Homogenization limits and Wigner transforms", Proposition 1.8) and the result is also implicit in (Ryzhik, Papanicolaou, and Keller, "Transport equations for elastic and other waves in random media").
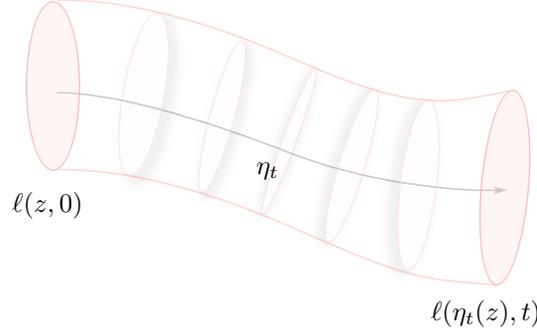
Figure 3.4: Transport of the phase space energy density along a ray.

3.2.2 The Light Transport Equation

The light transport equation in Eq. 3.39 is the time evolution equation for the phase space light energy density in a medium with inhomogeneous refractive index. It hence plays a central role in light transport theory. With the definition of the canonical Poisson bracket on the cotangent bundle the equation becomes

$$\frac{\partial \ell_t}{\partial t} = -\{\ell_t, H\} = -\frac{\partial \ell_t}{\partial q}\frac{\partial H}{\partial p} + \frac{\partial \ell_t}{\partial p}\frac{\partial H}{\partial q} = -\frac{\partial \ell_t}{\partial q}\frac{\partial q}{\partial t} - \frac{\partial \ell_t}{\partial p}\frac{\partial p}{\partial t} \qquad (3.49)$$

where the last equality follows from Hamilton's equations. A comparison with the discussion in Remark 2.117 shows that Eq. 3.49 is the material derivative

$$0 = \frac{d}{dt}\ell_t = \frac{\partial \ell_t}{\partial t} + \frac{\partial \ell_t}{\partial q}\frac{\partial q}{\partial t} + \frac{\partial \ell_t}{\partial p}\frac{\partial p}{\partial t} \qquad (3.50a)$$

of the light energy density. Using tensor notation, Eq. 3.50a can be written as

$$0 = \frac{d}{dt}\ell_t = \frac{\partial \ell_t}{\partial t} + \pounds_{X_H}\ell_t \qquad (3.50b)$$

which follows from considering divergence with respect to the Liouville form which is preserved along the flow of the Hamiltonian vector field X_H. By Eq. 3.50, the light transport equation expresses the conservation of the energy density ℓ_t along the flow η_t generated by the Hamiltonian vector field X_H. By the definition of the Lie derivative for time-varying tensor fields, cf. Remark 2.157, and using Eq. 3.50b, the conservation of the light energy density along trajectories in phase space can also be written using the pullback

$$\eta_t^* \ell_t = \ell_0 \qquad (3.51a)$$

along the Hamiltonian flow η_t, and it provides a concise description of the transport of ℓ_t. More explicitly, the pullback is written as

$$(\eta_t^* \ell_t)(q, p) = \ell_t(\eta_t(q, p)) = \ell_0(q, p), \qquad (3.51b)$$

cf. Fig. 3.4

The transport along the Hamiltonian flow η_t also preserves the angular frequency of the phase space light energy density so that for $\ell_0(q_0, p_0)$ and $\ell_t(q_t, p_t)$ where the phase space points are related by $(q_t, p_t) = \eta_t(q_0, p_0)$ we have $\omega_0 = \omega_t$. By the conservation of the Hamiltonian $H(q_t, p_t) = H(q_0, p_0)$ along the flow one immediately has

$$\frac{\|p_0\|}{\|p_t\|} = \frac{n(q_0)}{n(q_t)}. \tag{3.52a}$$

But from the definition of the momentum in Eq. 3.24 it also follows that

$$\frac{\|p_0\|}{\|p_t\|} = \frac{n(q_0)\,\omega_0}{n(q_t)\,\omega_t}. \tag{3.52b}$$

A comparison of Eq. 3.52a and Eq. 3.52b shows that we have to have $\omega_0 = \omega_t$ and frequency is conserved.

The description of the dynamics in Eq. 3.39 is equivalent to more classical formulations of the transport equation. With the coordinate expression of the Poisson bracket, cf. Eq. 3.49, and using ∇-notation we have

$$\frac{\partial \ell_t}{\partial t} = -\nabla_q \ell \cdot \frac{\partial \mathcal{H}}{\partial p} + \nabla_p \ell_t \cdot \frac{\partial \mathcal{H}}{\partial q} \tag{3.53a}$$

and using Hamilton's equations in Eq. 3.40 we obtain

$$\frac{\partial \ell_t}{\partial t} = -\frac{c}{n} \frac{p}{\|p\|} \cdot \nabla_q \ell_t + \frac{c\,\|p\|}{n^2} \nabla_q\, n \cdot \nabla_p \ell_t. \tag{3.53b}$$

Rearranging terms the yields

$$\frac{n}{c} \frac{\partial \ell_t}{\partial t} + \frac{p}{\|p\|} \cdot \nabla_q \ell_t - \frac{1}{n} \|p\| \nabla_q\, n \cdot \nabla_p \ell_t = 0 \tag{3.53c}$$

which is the form of the transport equation for the light energy density in a medium with isotropic but spatially varying refractive index that can be found in the literature.[99] The interpretation of the terms in Eq. 3.53c is again provided by the equivalence with the material derivative.

Remark 3.7. In applications, one is usually interested in steady state solutions. Pragmatically, this implies that the Eulerian rate of change vanishes $\partial \ell_t / \partial t$ vanishes although steady state systems have usually very special properties that can be characterized on a more intrinsic level.[100]

---

[99]See for example (Bal, "Radiative transfer equations with varying refractive index: a mathematical perspective").

[100]See for example (Marsden, Ratiu, and Abraham, *Manifolds, Tensor Analysis, and Applications*, Proposition 9.2.2) and (Arnold and Khesin, *Topological Methods in Hydrodynamics*, Chapter II) for characterizations of steady flows of the ideal Euler fluid, and (Golubitsky and Stewart, *The Symmetry Perspective: From Equilibrium to Chaos in Phase Space and Physical Space*) for some general results concerning steady state solutions.
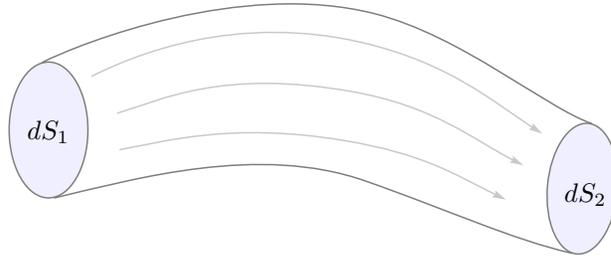
Figure 3.5: Geometry of the intensity law of geometric optics.

Remark 3.8. In classical optics, the light trajectories $r(s)$ in an inhomogeneous medium with varying refractive index are described by the eikonal equation[101]

$$\frac{d}{ds}\left(n(q)\frac{r(s)}{ds}\right) = \nabla_q n(q) \tag{3.54}$$

where ds denotes parametrization by arc length. The transport of energy is then described by the intensity law of geometric optics

$$I_1 \, dS_1 = I_2 \, dS_2 \tag{3.55}$$

where I_1 is the intensity at the beginning of an infinitesimal tube of rays with area dS_1 and I_2 is the intensity at the end of the tube with area dS_2, cf. Fig. 3.5.[102] The derivation of the intensity law is rather different than those of the light transport equation presented in the foregoing in Chapter 3.2.1. Nonetheless, the conservation of intensity I during transport suggests that it might be related to the phase space light energy density ℓ.

From Fermat's principle it follows that $dt = n(q)/c \, ds$ and hence the eikonal equation in Eq. 3.54 can also be written as

$$\frac{d}{dt}\dot{q} = \frac{c}{n}\nabla_q n. \tag{3.56}$$

Assuming the intensity to be a density $e = I(q, \dot{q}) \, dq \wedge dp \in \mathrm{Den}(TQ)$ over velocity phase space $T^Q \cong T^*Q$, which we identify with T^*Q using the canonical metric on $\mathbb{R}^3$, its material derivative is given by

$$0 = \frac{d}{dt}e = \frac{\partial e}{\partial t} + \frac{\partial e}{\partial q}\frac{\partial q}{\partial t} + \frac{\partial e}{\partial \dot{q}}\frac{\partial \dot{q}}{\partial t} \tag{3.57}$$

---

[101] Born and Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Chapter 3.2.1, with $dt = n(q)ds$.

[102] The intensity law is based on the divergence freeness of the Poynting vector in a non-conducting medium (ibid., Chapter 1.4.3,Chapter 3.1.2). In previous work in computer graphics it was considered by (Ihrke et al., "Eikonal Rendering: Efficient Light Transport in Refractive Objects") but to our knowledge not by (Stam and Languénou, "Ray Tracing in Non-Constant Media").

and it vanishes by the intensity law of geometric optics. With Eq. 3.56 and using ∇-notation, we obtain for the material derivative

$$0 = \frac{\partial e}{\partial t} + \frac{c}{n} \frac{\dot{q}}{\|\dot{q}\|} \cdot \nabla_q \, e + \frac{1}{n} \nabla_q \, n \cdot \nabla_{\dot{q}} \, e \tag{3.58}$$

and where we also used the chain rule and $dt = n(q)/c \, ds$ to simplify $\partial q / \partial s$. This is equivalent to

$$0 = \frac{n}{c} \frac{\partial e}{\partial t} + \frac{\dot{q}}{\|\dot{q}\|} \cdot \nabla_q \, e + \frac{1}{c} \nabla_q \, n \cdot \nabla_{\dot{q}} \, e. \tag{3.59}$$

To relate the last equation to the transport equation for the phase space energy density in Eq. 3.53c, we use the identity

$$\frac{\|p\|}{\|p\|} = \frac{c}{n} \frac{1}{\omega} \|p\| \tag{3.60}$$

and choose $\omega = 1$, which is possible without loss of generality since the frequency in geometric optics is arbitrary, see also the subsequent discussion of the contact structure of light transport in Chapter 3.2.4. We therefore obtain

$$0 = \frac{n}{c} \frac{\partial e}{\partial t} + \frac{\dot{q}}{\|\dot{q}\|} \cdot \nabla_q \, e + \frac{1}{c} \frac{c}{n} \|p\| \, \nabla_q \, n \cdot \nabla_{\dot{q}} \, e. \tag{3.61}$$

The time evolution equation for the intensity density e is hence given by

$$0 = \frac{n}{c} \frac{\partial e}{\partial t} + \frac{\dot{q}}{\|\dot{q}\|} \cdot \nabla_q \, e + \frac{1}{n} \|p\| \nabla_q \, n \cdot \nabla_{\dot{q}} \, e. \tag{3.62}$$

and it is equivalent to the transport equation for the phase space light energy density ℓ in Eq. 3.53c when e is identified with ℓ and the velocity $\dot{q}$ with the momentum p using the canonical metric on Euclidean space.[103]

Remark 3.9. As remarked before, the light transport equation is in the form of the Vlasov or collisionless Boltzmann equation[104] which arose in the kinetic theory of gases to describe the time evolution of a density of gas molecules $f \in \mathrm{Den}(T^*Q)$.[105] Classically, the equation is justified by the Born-Bogoliubov-Green-Kirkwood-Yvon hierarchy (BBGKY)[106] which provides a rigorous way

---

[103]In the literature the shading equation for (pre-)radiance and not the light energy density was used together with intensity, see for example (Ihrke et al., "Eikonal Rendering: Efficient Light Transport in Refractive Objects"), which appears to be inconsistent with the above result also will become clear from the discussions in Chapter 3.2.5 and Chapter 3.2.7.

[104]In the computer graphics literature this has been pointed out before by Arvo ("Transfer Equations in Global Illumination") although it is a well known fact in general transport theory (Duderstadt and Martin, *Transport Theory*).

[105]Kinetic theory originates in classical work by Maxwell ("On the Dynamical Theory of Gases") and Boltzmann (*Vorlesungen über Gastheorie*) and gases. Modern kinetic theory is concerned with systems compromised of a large number of similar or identical objects, see for example (Liboff, *Kinetic Theory: Classical, Quantum, and Relativistic Descriptions*).

[106]Interestingly, the BBGKY hierarchy has a natural geometric interpretation as shown in (Marsden, Morrison, and Weinstein, "The Hamiltonian structure of the BBGKY hierarchy equations"). For an overview of variations of and alternatives to the BBGKY hierarchy see (Duderstadt and Martin, *Transport Theory*, p. 35).

to reduce the Liouville equation

$$\frac{\partial \rho_n}{\partial t} + \{\rho_n, H_n\} = 0 \tag{3.63}$$

to the Boltzmann equation. The Liouville equation describes the time evolution of an n-particle distribution

$$\rho_n(z_1, z_2, \ldots, z_n)\, dz^1 \ldots dz^n \in \mathrm{Den}\left((T^*Q)^n\right), \tag{3.64}$$

representing the probability that the first particle is at $z_1 = (q_1, p_1)$, the second one at $z_2 = (q_2, p_2)$, and so forth, for some suitable n-particle Hamiltonian H_n describing the joined dynamics of the system.[107] Liouville's theorem thereby asserts that the n-particle phase space distribution ρ_n is conserved under the dynamics. The number n of particles is typically very large, in gas theory for example of the order of the Avagadro number 6.022×10^{23}, so that the system is still well described when only k particles are considered explicitly and one averages over the remaining ones by

$$\rho_k(z_1, \cdots, z_k) = \int \cdots \int \rho(z_1, \cdots, z_k, z_{k+1}, \cdots, z_n)\, dz^{k+1} \cdots dz^n \tag{3.65}$$

Associated to the kth-order BBGKY moment distributions $\rho_k \in \mathrm{Den}\left((T^*Q)^k\right)$ is an evolution equation with a kth-order Hamiltonian. By integrating out all but one particle one obtains the single particle phase space distribution

$$f(q, p) \equiv \rho_1(q_1, p_1) \in \mathrm{Den}(T^*Q) \tag{3.66}$$

and its dynamics are described by the Boltzmann equation

$$\frac{\partial f}{\partial t} + \{f, H\} = \left(\frac{\partial f}{\partial t}\right)_{\text{scatt.}} \tag{3.67}$$

where the right hand side models collisions between the gas molecules, for example with Boltzmann's famous Stoßzahlansatz, and it is meant to compensate for some of the effects that have been neglected by using only the single particle distribution $f = \rho_1$. The Vlasov or collisionless Boltzmann equation is obtained when the right hand side of Eq. 3.67 is neglected and the equation occurs naturally in various disciplines.[108] The single particle distribution f then represents for example a density over electrons or galaxies and the transported particles usually interact with a field such as an electromagnetic field in the Maxwell-Vlaslov and Vlasov-Poisson systems in plasma physics or the gravitational field in applications in astrophysics. Systems in kinetic theory

---

[107]The n-particle distribution ρ_n can either be interpreted as a density or as a probability distribution. Both viewpoints are fully equivalent and lead to identical results (Villani, "A Review of Mathematical Topics in Collisional Kinetic Theory").

[108]See (Henon, "Vlasov Equation") for some historical remarks on the Vlasov equation. Physical contexts where the equation arises are discussed in (Holm, Putkaradze, and Tronci, "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions").

have recently also been studied from a geometric point of view[109] and the rich structure obtained in this work provided a significant motivation for the group structure of ideal light transport which will be discussed in Chapter 3.3. Light transport theory can be considered as part of a generalized kinetic theory since the governing equations have the same structure but, as discussed in detail in Chapter 3.2.1, its physical justification derives from electromagnetic theory and not the BBGKY hierarchy. Additionally, for light transport scattering with the transport medium and at boundaries is significantly more important than for most other systems in kinetic theory.

Remark 3.10. The Boltzmann and the Vlasov equation represent a first-order approximation within the BBGKY hierarchy. However, they are still rather complicated with dynamics on a $2n$-dimensional phase space. Statistical and kinetic moments, which were first systematically studied by Chapman and Enksog,[110] enable to obtain simplified models from the Boltzmann and Vlasov equation that are more directly related to observables. Kinetic moments are defined as

$$A_n = \int_{T_q^* Q} f(q,p)\, p^n \, dp \tag{3.68a}$$

while statistical moments are taken with respect to position and momentum and hence given by

$$B_{n,m} = \int_{T^* Q} f(q,p)\, q^n\, p^m \, dq\, dp. \tag{3.68b}$$

Taking moments is a Poisson map from the space of smooth functions under the Poisson bracket to the space of moment coefficients,[111] cf. Remark 2.156, which intuitively follows when the moments are considered as basis coefficients for a Taylor or polynomial basis expansion and the bilinearity of the Poisson bracket is exploited. Kinetic moments are continuous in configuration space and this

---

[109]For the Maxwell-Vlasov and Vlasov-Poisson system see (Marsden and Weinstein, "The Hamiltonian Structure of the Maxwell-Vlasov equations"; Marsden et al., "Hamiltonian Systems with Symmetry, Coadjoint Orbits and Plasma Physics"; Marsden, "A Group Theoretic Approach to the Equations of Plasma Physics"; Cendra et al., "The Maxwell-Vlasov Equations in Euler-Poincaré Form"; Gümral, "Geometry of plasma dynamics. I. Group of canonical diffeomorphisms"). The Hamiltonian structure of the Jeans equation in astrophysics was studied by Kandrup ("The secular instability of axisymmetric collisionless star clusters"). More recently, also general kinetic equations have been studied, see (Holm, Putkaradze, and Tronci, "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions") and references therein.

[110]Chapman and Cowling, *The Mathematical Theory of Non-uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases*.

[111]For kinetic moments the Lie algebra homomorphism was studied in detail in (Gibbons, Holm, and Tronci, "Geometry of Vlasov kinetic moments: A bosonic Fock space for the symmetric Schouten bracket") and it was shown that the Lie bracket for the space of moments is the Schouten bracket. For statistical moments the structure was investigated in earlier work (Gibbons, "Collisionless Boltzmann Equations and Integrable Moment Equations"; Channell, "The beam-beam interaction as a discrete Lie-Poisson dynamical system"; Holm, Lysenko, and Scovel, "Moment invariants for the Vlasov equation"; Channell, "Canonical Integration of the Collisionless Boltzmann Equation"; Lysenko and Parsa, "Beam Matching and Halo Control").

enables to obtain reduced continuous systems modeled by differential equations. For example, the Euler and Navier-Stokes equations can be deduced from the Boltzmann equation by considering only the zero and first order moment of the phase space density, representing the fluid density and average velocity, respectively,[112] and this transition is analogous to the equivalence of the light transport equation and Poynting's theorem in Remark 3.6, see also Eq. 3.1 and Eq. 3.4. Other macroscopic models for fluid flow such as Darcy's law can also be derived as an approximation to the Vlasov equation, and measurable quantities correspond then again in many cases to moments.[113] Statistical moments are fully discrete, which can be exploited to obtain finite representations for numerical computations.[114]

3.2.3 The Hamiltonian Structure of Light Transport

In this section, we will consider the Hamiltonian structure of light transport in more detail. We will employ a Legendre transform to obtain the Hamiltonian for light transport in Eq. 3.38 from Fermat's principle, and the derivation will show that light transport is a geodesic flow for a metric defined by the refractive index, see Fig. 3.6.[115]

Fermat's principle states that light takes paths of extremal time, that is optical paths are stationary points

$$\delta \int_0^T dt = 0 \tag{3.69}$$

of the time functional. Changing to a parametrization by arc length so that $\|\dot{q}\|^2 = \delta_{ij}\dot{q}^i\dot{q}^j = 1$ we have

$$\delta \int_a^b \frac{1}{\|\dot{q}\|} ds = \delta \int_a^b \frac{n(q)}{c} ds = 0 \tag{3.70}$$

since the speed of light in a medium is $c/n(q)$. As is well known, Eq. 3.70 can be interpreted as a stationary point of the length functional

$$\frac{\delta\ell(q)}{\delta q} = \delta \int \|\dot{q}\|_n ds = \delta \int \sqrt{g_{ij}^n \dot{q}^i\dot{q}^j} ds = 0, \tag{3.71}$$

[112]The mapping provided by the moments is a momentum map known as plasma-to-fluid map and it was discovered in (Marsden et al., "Hamiltonian Systems with Symmetry, Coadjoint Orbits and Plasma Physics"); see also (Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, p. 367).

[113]See (Holm, Putkaradze, and Tronci, "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions"; Tronci, "Geometric dynamics of Vlasov kinetic theory and its moments") and references therein.

[114]Preliminary work that employs statistical moments for computer simulations can be found in (Holm, Lysenko, and Scovel, "Moment invariants for the Vlasov equation"; Scovel and Weinstein, "Finite Dimensional Lie-Poisson Approximations to Vlasov-Poisson Equations"; Channell, "Canonical Integration of the Collisionless Boltzmann Equation").

[115]Much of the following material, in particular those related to the geodesic description of light transport, can be found in similar form in (Novikov and Taimanov, *Modern Geometric Structures and Fields*, p. 479, 514).

geodesic energy $\quad \hat{L} = n^2(q) \quad \xleftrightarrow[\text{transform}]{\text{Legendre}} \quad \hat{H} = g^{ij}\hat{p}_i\hat{p}_j \quad$ geodesic Hamiltonian

Fermat's principle $\quad L = n(q) \qquad\qquad\qquad \sqrt{\hat{H}} = \frac{c}{n(q)}\|\hat{p}\|$

$L(q, \dot{q}) = \omega\sqrt{g_{ij}^n \dot{q}^i\,\dot{q}^j} \qquad\qquad H = \omega\check{H} = \frac{c}{n(q)}\|p\| \quad$ light transport
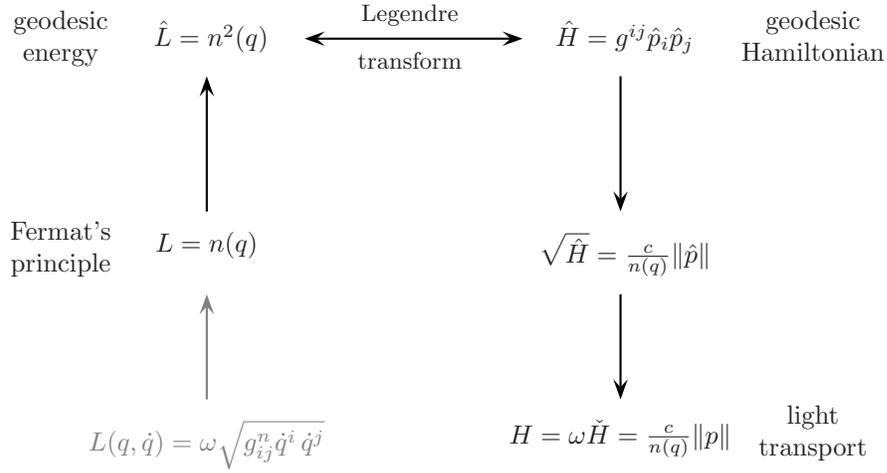
Figure 3.6: Schematic view of the Legendre transform from Fermat's principle to light transport.

cf. Definition 2.167. Hence, the natural metric $g_n \in \mathcal{T}_0^2(Q)$ for light transport is given by

$$g_{ij}^n = \begin{pmatrix} n^2(q)/c^2 & 0 & 0 \\ 0 & n^2(q)/c^2 & 0 \\ 0 & 0 & n^2(q)/c^2 \end{pmatrix} \tag{3.72a}$$

and the associated co-metric $g^n \in \mathcal{T}_2^0(Q)$ has the form

$$g_n^{ij} = \begin{pmatrix} c^2/n^2(q) & 0 & 0 \\ 0 & c^2/n^2(q) & 0 \\ 0 & 0 & c^2/n^2(q) \end{pmatrix} \tag{3.72b}$$

or more concisely $g_{ij}^n = n^2(q)/c^2\,\delta_{ij}$ and $g_n^{ij} = c^2/n^2(q)\,\delta^{ij}$.[116] Fermat's principle can hence also be interpreted as defining light rays as (local) geodesics with respect to the metric g_n defined by the refractive index $n(q)$.

One can regard Eq. 3.71 as an action integral with Lagrangian

$$L(q, \dot{q}) = \sqrt{g_{ij}^n \dot{q}^i\dot{q}^j} \tag{3.73}$$

but the Legendre transform is then degenerate and one obtains a vanishing Hamiltonian $H(q, p) = 0$.[117] Hence, using Proposition 2.77, instead of the length

---

[116]The metric g_n is closely related to the natural metric for the electromagnetic field in Eq. 3.14 and it appeared to our knowledge first in (Gordon, "Zur Lichtfortpflanzung nach der Relativitätstheorie").

[117]Kline and Kay, *Electromagnetic Theory and Geometrical Optics*.

we will consider the energy functional[118]

$$\frac{\delta E(q)}{\delta q} = \delta \int \|\dot{q}^i\|_n^2 \, ds = \delta \int g_{ij}^n \, \dot{q}^i \dot{q}^j \, ds = 0 \tag{3.74}$$

which is an action functional for the Lagrangian

$$\hat{L}(q, \dot{q}) = g_{ij}^n \, \dot{q}^i \dot{q}^j. \tag{3.75}$$

Following the usual ansatz for the Legendre transform, as discussed in Chapter 2.3.4.2, the canonical momentum $\hat{p}$ is

$$\hat{p}_k = \frac{\partial \hat{L}}{\partial \dot{q}_k} = g_{ij}^n \frac{\dot{q}^i}{\dot{q}^k} \dot{q}^j + g_{ij}^n \dot{q}^i \frac{\dot{q}^j}{\dot{q}^k} = g_{ij}^n \delta^{ik} \dot{q}^j + g_{ij}^n \dot{q}^i \delta^{jk} = 2 g_{kj}^n \dot{q}^j \tag{3.76}$$

and it is thus obtained by lowering the indices of the velocity $\dot{q}$ with respect to the metric g_n.

Remark 3.11. The canonical momentum,

$$\hat{p}_k = 2 g_{kj}^n \dot{q}^j = 2 \frac{n^2(q)}{c^2} \delta_{ik} \left(\frac{c}{n(q)} \bar{s}^i \right) = 2 \frac{n(q)}{c} \delta_{ik} \bar{s}^i = 2 \frac{n(q)}{c} \bar{s}_k, \tag{3.77}$$

has to be distinguished from the kinetic momentum,

$$p_k = \vec{k}^\flat = \frac{n(q)}{c} \, \omega \, \bar{s}_k, \tag{3.78}$$

that is defined by the wave vector $\vec{k}$ and more closely related to energy transport.[119] The angular frequency factor $\omega = 2\pi\nu$ in which Eq. 3.77 and Eq. 3.78 differ can be associated with an $\mathbb{R}^+$ symmetry in the fibers and $\hat{p}$ can thus be considered as the reduced momentum on the cosphere bundle S^*Q,[120] as will be discussed in more detail in the next section when we study the contact structure of light transport.

With the definition of the Hamiltonian in terms of the Lagrangian in Def. 2.206 and with the expression for the canonical momentum in Eq. 3.76, we obtain

$$\hat{H} = \hat{p}_i q^i - L = 2 g_{ij}^n \dot{q}^i \dot{q}^j - g_{ij}^n \dot{q}^i \dot{q}^j = g_{ij}^n \dot{q}^i \dot{q}^j = \hat{p}_i(\dot{q}^i). \tag{3.79}$$

By Eq. 3.24, the velocity $\dot{q}$ is given by

$$\dot{q}^i = g^{ij} \hat{p}_j = \frac{c^2}{n^2(q)} \delta^{ij} \hat{p}_j \tag{3.80}$$

---

[118]The idea to employ the energy functional instead of the length functional for the Legendre transform is from the second printing of the second edition of the book by Arnold (*Mathematical Methods of Classical Mechanics*).

[119]A similar distinction between a canonical and kinetic momentum exists in electromagnetic theory, see for example (Jackson, *Classical Electrodynamics*, Chapter 12), and our terminology stems from this analogy.

[120]The inessential factor of 2 in Eq. 3.77 could be avoided by defining the Lagrangian with an additional factor of $1/2$ as is commonly done for the kinetic energy.

and hence we have

$$\hat{H} = \hat{p}_i \left(\frac{c^2}{n^2(q)} \delta^{ij} \hat{p}_j \right) = \frac{c^2}{n^2(q)} \delta^{ij} \hat{p}_i \hat{p}_j = \frac{c^2}{n^2(q)} \|\hat{p}\|^2 = g^{ij} \hat{p}_i \hat{p}_j \qquad (3.81)$$

which is the well known expression for a geodesic Hamiltonian.[121] The Hamilton $H(q, p)$ for light transport in Eq. 3.38 is obtained as

$$H(q, p) = \omega \sqrt{\hat{H}(q, p)} = \frac{c}{n(q)} \omega \|\hat{p}\|. \qquad (3.82)$$

and the Hamiltonian vector field X_H for light transport can thus be written as

$$X_H = \omega \frac{1}{\sqrt{H_g(q, p)}} X_{\hat{H}}. \qquad (3.83)$$

The vector fields only differ by a scalar factor in each fiber, which can be interpreted as a time dilation, and hence their flows coincide. Intuitively, the transition from $\hat{H}$ to H can be seen as reverting the transition from the length to the energy functional in Eq. 3.71 and Eq. 3.74, and then multiplying by the angular frequency ω to "lift" from S^*Q to T^*Q, see again Fig. 3.6 and also the discussion on the contact structure in the next section.

Remark 3.12. The need to employ the squared Lagrangian of the original system for the Legendre transform and the intrinsic connection to a geodesic flow which we developed can be considered as an instance of a general theory where truncated actions for a fixed value of the Hamiltonian are employed,[122] which for our system corresponds to a fixed value of the angular frequency ω.

Remark 3.13. Mathematically the justification for working with the energy instead of the length functional is that the former is considerably better behaved and that, under reasonable assumptions, their extrema coincide.[123]

Remark 3.14. It is apparent from Fig. 3.6 that one could define a Lagrangian $L(q, \dot{q}) = \omega \sqrt{g^n_{ij} \dot{q}^i \dot{q}^j} = \omega n(q)$. This would make the diagram symmetric and ω could be retained as a constant factor throughout the Legendre transform.

Remark 3.15. For the construction of the geodesic flow that was obtained in the foregoing, it is sufficient to have a Finsler metric and no Riemannian structure is required.[124]

Remark 3.16. Paraxial optics has been studied extensively from the point of view of Hamiltonian mechanics, and symmetries associated with radial and translation invariance of the refractive along the optical axis have been

---

[121] Jost, *Riemannian Geometry and Geometric Analysis*, Theorem 1.9.3.

[122] Novikov and Taimanov, *Modern Geometric Structures and Fields*, Chapter 12.4.3.

[123] See for example (Castro and Tromba, *Lecture Notes on Morse Theory*).

[124] See for example (Pettini, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Chapter 3.2).

studied.[125] Such symmetries do not exist for the general setting we are concerned with and for us the energy density on phase space is central, which also has not been considered in previous work on paraxial optics.

3.2.4 The Contact Structure of Light Transport[126]

Next to the Hamiltonian structure considered in the previous sections, light transport is also endowed with a natural contact structure that arises from an $\mathbb{R}^+$ symmetry of the Hamiltonian. The symmetry enables to reduce the dynamics of trajectories from the slit cotangent bundle $T^*Q \setminus \{0\}$ to the five dimensional cosphere bundle S^*Q, although due to the frequency dependence no such reduction is easily possible for the dynamics of the phase space light energy density.

In Eq. 3.52, we established conservation of frequency under the time evolution of the light energy density. By Noether's theorem this implies the existence of a continuous symmetry and a reduced geometric structure. The symmetry arises from the light transport Hamiltonian which is positive homogeneous of degree one in the momentum and satisfies

$$H(q, \gamma\, p) = \frac{c}{n(q)}\, \|\gamma\, p\| = \gamma\, \frac{c}{n(q)}\, \|\, p\| = \gamma\, H(q, p) \;\; , \;\; \gamma \in \mathbb{R}^+ \tag{3.84}$$

and the strict positivity of $\gamma \in \mathbb{R}^+$ follows from the momentum being proportional to the frequency[127] and since $c/n(q)$ is positive by $\dot{q} = c/n(q)$.[128] The strict positivity implies that it is sufficient to consider the slit cotangent bundle $T^*Q \setminus \{0\}$ as phase space. With the $\mathbb{R}^+$ symmetry, the reduced phase space is the cosphere bundle

$$S^*Q = (T^*Q \setminus \{0\}) / \mathbb{R}^+ \tag{3.85}$$

where all momenta p in a radial section in a fiber $T_q^*Q \setminus \{0\}$ are identified as a coset

$$[p] = \left\{ p \in T_q^*Q \;\middle|\; p = \gamma \frac{\bar{p}}{\|\bar{p}\|}, \gamma \in \mathbb{R}^+ \right\} \in S_q^*Q, \tag{3.86}$$

see Fig. 3.7. Moreover, by Theorem 2.34 also the dynamics project onto the cosphere bundle S^*Q and the symplectic flow $\eta_t^\theta : \mathbb{R} \times T^*Q \setminus \{0\} \to T^*Q \setminus \{0\}$

[125]Holm and Wolf, "Lie-Poisson Description of Hamiltonian Ray Optics"; Rivera, Chumakov, and Wolf, "Hamiltonian foundation of geometrical anisotropic optics"; Wolf, *Geometric Optics on Phase Space*; Holm, *Geometric Mechanics: Dynamics and Symmetry*.

[126]For the necessary mathematical background on contact structures and the cosphere bundle the reader is referred to Chapter 2.3.4.5.

[127]Excluding the zero section is also necessary to avoid an infinite wavelength which would otherwise require special treatment in various contexts.

[128]Recently, materials with negative refractive index have been developed, which are so called meta-materials with properties not found in nature, see for example (Cui et al., *Metamaterials: theory, design, and applications*) and references therein. We will exclude these exotic materials from our considerations.
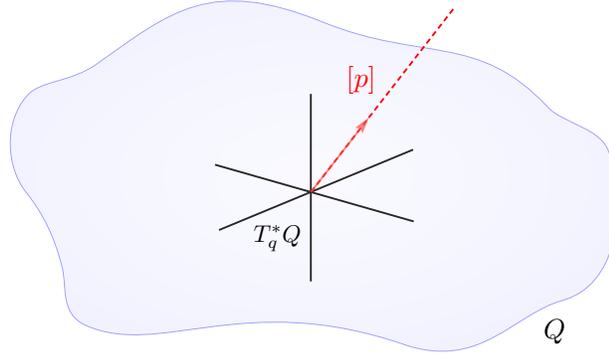
Figure 3.7: Cosets $[p]$ defined in fibers T_q^*Q through the $\mathbb{R}^+$ symmetry of light transport. With the natural Riemannian metric g_n for light transport, each coset can be represented by the unit covectors $\bar{p}^n$.

of the Hamiltonian system becomes a contact diffeomorphism $\tilde{\eta}_t : \mathbb{R} \times S^*Q \to S^*Q$, and this diffeomorphism commutes with the multiplicative $\mathbb{R}^+$-action $m_\gamma(q, p) = (q, \gamma p)$. Hence, the reduced dynamics are given by[129]

$$
\begin{array}{ccc}
T^*Q \backslash \{0\} & \xrightarrow{\eta_t^\theta} & T^*Q \backslash \{0\} \\
m_\gamma \uparrow & & \uparrow m_\gamma \\
S^*Q & \xrightarrow[\tilde{\eta}_t]{} & S^*Q
\end{array}
\tag{3.87}
$$

and time evolution can be determined on the five dimensional cosphere bundle S^*Q in which case the frequency can be reconstructed from a representative element in each coset $[p]$ using the multiplicative m_γ action. The diagram in Eq. 3.87 also implies that the cosets $[p]$ are preserved under the dynamics, see Fig. 3.8, and, in fact, this is already imprinted in Hamilton's equations for light transport in Eq. 3.40. The velocity for light transport is given by

$$
\dot{q} = \frac{\partial H}{\partial p} = \frac{c}{n(q)} \frac{p}{\|p\|}
\tag{3.88}
$$

and since $\dot{q}$ does only depend on the normalized momentum $p/\|p\|$ the cosets $[p]$ are trivially preserved. The momentum change

$$
\dot{p} = -\frac{\partial H}{\partial q} = \frac{c}{n^2(q)} \|p\| \nabla_n(q)
\tag{3.89}
$$

consists of a fiber invariant part

$$
\dot{p}_q = \frac{c}{n^2(q)} \nabla_n(q)
\tag{3.90}
$$

---

[129]For clarity, the inclusion map $i : S^*Q \to T^*Q$ has been omitted in the commutative diagram in Eq. 3.87.
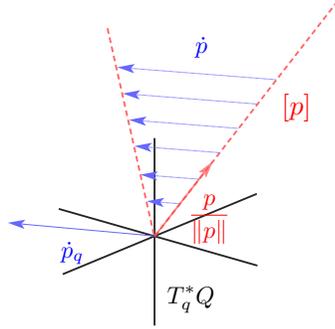
Figure 3.8: Effect of the Hamiltonian vector field for light transport on the fiber coordinate. Cosets $[p]$ are preserved since $\dot{p}$ can be factored into a fiber invariant part $\dot{p}_q$ and a radial component $\|p\|$ so that $\dot{p} = \dot{p}_q \, \|p\|$.

and a radial component $\|p\|$. Elementary trigonometry shows that $\|p\|$ leads to the radial scaling of $\dot{p}_q$ that is needed to preserve the cosets $[p]$ under the dynamics, cf. Fig. 3.8.[130] Hence, also from the Hamiltonian vector field we can conclude that for the dynamics it is sufficient to consider an arbitrary section in $T^*Q \backslash \{0\}$, representing the cosets $[p]$, and that a complete radial section can be reconstructed using the m_γ action.

A natural choice for the section in the slit cotangent bundle $T^*Q \backslash \{0\}$ is provided by the unit section with respect to the co-metric g^n for light transport in Eq. 3.72. With $\bar{q}$ being a unit vector for the canonical Euclidean metric, a unit vector with respect to the metric g_n is

$$\bar{q}_n = \frac{c}{n(q)} \bar{q} \tag{3.91}$$

since we then have

$$g^n_{ij} \, \bar{q}^i_n \bar{q}^j_n = \frac{n^2(q)}{c} \delta_{ij} \left(\frac{c}{n(q)} \bar{q}^i \right) \left(\frac{c}{n(q)} \bar{q}^j \right) = \frac{n^2(q)}{c^2} \frac{c^2}{n^2(q)} \, \delta_{ij} \, \bar{q}^i \bar{q}^j = 1. \tag{3.92}$$

A unit covector $\bar{p}^n = (\bar{p}^n_1, \bar{p}^n_2, \bar{p}^n_3)$ for g^n can now be obtained by lowering indices and it is thus given by

$$\bar{p}^n_j = g^n_{ij} q^i = \frac{n^2(q)}{c^2} \delta_{ij} \, \frac{c}{n(q)} \bar{q}^i = \frac{n(q)}{c} \bar{q}_j. \tag{3.93}$$

The space of all such covectors $\bar{p}_n$ forms the Riemannian cosphere bundle

$$S^*_n Q = \left\{ p \in T^*Q \backslash \{0\} \, \middle| \, p_j = \frac{n(q)}{c} \bar{q}_j \right\} \tag{3.94}$$

on the Riemannian manifold (Q, g_n), and each unit covector $\bar{p}^n$ represents a respective coset $[p]$, see again Fig. 3.7. The contact 1-form for $S^*_n Q$ is hencethe

---

[130]The result follows immediately by considering $\tan(\alpha) = \|\dot{p}_q\| = \|\dot{p}\|/\|p\|$.

restriction of the canonical 1-form $\theta = p_i dq^i$ on $T^*Q\backslash\{0\}$ to the cosphere bundle given by

$$\theta_n = \theta \mid_{S_n^*Q} \tag{3.95}$$

and using θ_n we can identify the contact vector field X_n and the contact Hamiltonian H_n for the diffeomorphism $\tilde{\eta}_t : \mathbb{R} \times S^*Q \to S^*Q$ obtained from the Hamiltonian flow η_t^θ on $T^*Q\backslash\{0\}$. The existence of X_n and H_n is assured by Theorem 2.34, and from this theorem we also know that

$$H_n = \theta_n(X_n). \tag{3.96}$$

For light transport the contact Hamiltonian H_n is the canonical Hamiltonian H in Eq. 3.38 but with the momentum restricted to the cosphere bundle S_n^*Q in Eq. 3.94. Hence, also X_n is the restriction of the canonical Hamiltonian vector field X_H to S_n^*Q. Indeed, applying Eq. 3.96 to the Hamiltonian dynamics on $T^*Q\backslash\{0\}$ we have

$$\theta(X_H)(q,p) = p_i \, dq^i \left(\frac{c}{n(q)} \frac{p^i}{\|p\|} \frac{\partial}{\partial q^i} + \frac{c}{n^2(q)} \|p\| \nabla_q n(q) \frac{\partial}{\partial p^i} \right) \tag{3.97a}$$

and since the canonical 1-form only has support in the configuration space basis functions $\partial/\partial q^i$ but not in the fiber ones $\partial/\partial p^i$ we have

$$\theta(X_H)(q,p) = p_i \, \frac{c}{n(q)} \frac{p^i}{\|p\|} = \frac{c}{n(q)} \frac{p_i \, p^i}{\|p\|} = \frac{c}{n(q)} \frac{\|p\|^2}{\|p\|} = \frac{c}{n(q)} \|p\| \tag{3.97b}$$

which is the canonical Hamiltonian, that is $H = \theta(X_H)$. Hence, Eq. 3.96 is indeed satisfied when the restrictions θ_n and X_n are employed.

Next to the contact Hamiltonian, an alternative characterization of contact dynamics was the Reeb vector field R introduced in Def. 2.217, and we discussed in the foregoing that it is of unit length and the generator of the geodesic flow when the cosphere bundle is constructed on a Riemannian manifold, as it is the case for us. From Def. 2.217, we know that the Reeb vector field R has to satisfy

$$i_R \, \theta_n = \theta_n(R) = 1 \tag{3.98}$$

and since the Hamiltonian vector field X_H is the generator of the geodesic flow, which was shown in the last section, we have to have $R = X_n$. Indeed, with our previous results in Eq. 3.97 we obtain

$$\theta_n(X_n) = \frac{c}{n(q)} \left\| \frac{n(q)}{c} \, \bar{q} \right\| = \frac{c}{n(q)} \frac{n(q)}{c} = 1 \tag{3.99}$$

The second property that has to be satisfied by the Reeb vector field by Def. 2.217, $i_R \, d\theta_n = 0$, follows from a general result in the literature.[131]

---

[131]Ratiu and Schmid, "The differentiable structure of three remarkable diffeomorphism groups", p. 96, Remark 4.

Although the contact structure allows to reduce the Hamiltonian dynamics to the cosphere bundle, the following restriction is important to keep in mind: The light energy density ℓ is an arbitrary density on the (slit) cotangent bundle and a value $\ell(q, \bar{p}^n)$ for a representative momentum $\bar{p}^n$ does **not** provide information about $\ell(q, p)$ at any other value $p \in [p]$. Hence, $\ell(q, p)$ cannot be reconstructed when only $\ell(q, \bar{p}^n)$ is known, and the time evolution of the phase space light energy density cannot be projected onto the cosphere bundle S^*Q. As mentioned before, the question when the continuous function $\ell(q, p)$ can be reconstructed from "samples" $\ell(q, \bar{p})$ for selected values $\bar{p} \in [p]$ provides a central motivation for the work presented in the following chapter, and it will be discussed in detail there.

Remark 3.17. [132] The homogeneous Hamiltonian structure of light transport discussed in the present section enables to show the equivalence of our formulation to Huygens' principle, which describes light propagation as the superposition of an infinite number of spherical waves. Recall from Chapter 2.3.4.3 that a Lagrangian surface L_ω in a $2n$-dimensional symplectic manifold is an n-dimensional submanifold such that restriction of the symplectic 2-form ω to L_ω vanishes. A more restrictive notion is those of a conic Lagrangian surface L_θ for which the restriction of the canonical 1-form $\theta = p_i dq^i$ to the surface L_θ vanishes. An example for such a surface is provided by the cosphere bundle S_q^*Q, considered as a submanifold of the fiber $T_q^*Q \backslash \{0\}$, since the canonical 1-form has only support in the configuration space basis functions dq^i. With Remark 2.166, it is not hard to see that the symplectic flow $\eta_t^\theta : \mathbb{R} \times T^*Q \setminus \{0\} \to T^*Q \setminus \{0\}$ on $T^*Q \setminus \{0\}$ generated by a Hamiltonian homogeneous of degree one preserves a conic Lagrangian surface L_θ^0 unter its time evolution, that is, $L_\theta^t = \eta_t^\theta(L_\theta^0)$ is again a conic Lagrangian surface. Defining a wavefront at a time t as the projection of the conic Lagrangian surface

$$S_q^*Q_t = \eta_t^\theta(S_q^*Q_0) \tag{3.100}$$

onto configuration space, the equivalence of our formulation of light transport and Huygens' principle is obtained when the wavefront projections for all surfaces $S_q^*Q_t$ for which $q \in \mathcal{W}_0$ was on some initial wavefront $\mathcal{W}_0$ at time $t = 0$ are considered.

Remark 3.18. We believe that the contact structure and the reduced phase space provided by the cosphere bundle arise directly from the principal symbol of the Maxwell operator that is naturally defined on S^*Q. However, more work is required to verify this conjecture.

3.2.5 Phase Space Light Energy Density

The phase space light energy density ℓ, which provides a representation of the electromagnetic energy density on phase space, is *the* central quantity in light

---

[132]See for example (Geiges, *An Introduction to Contact Topology*, Chapter 1.5) or (Novikov and Taimanov, *Modern Geometric Structures and Fields*, p. 515).

transport theory. In the foregoing, we usually omitted its support in the time domain but a more comprehensive definition is

$$\ell_t = \ell(q,p,t) = \mathcal{L}(q,p,t)\, dq \wedge dp \wedge dt \in \mathrm{Den}(T^*Q \setminus \{0\} \times \mathbb{R}_t) \qquad (3.101)$$

since measurements of radiative energy always involve finite time intervals. In the above definition, we also employed the slit cotangent bundle $T^*Q \setminus \{0\}$ instead of canonical phase space T^*Q since there is no transport of energy when the frequency vanishes, and $T^*Q \setminus \{0\}$ will also prove mathematically convenient in the following. By Remark 2.150, the physical units of the energy density ℓ_t are those of energy, for example Joule, and we also have to require that ℓ has compact support, cf. Remark 2.120, so that its integral, representing energy, is guaranteed to be finite.[133]

Remark 3.19. The light energy density ℓ is a 1-density in the sense of Def. 2.146 and not as a volume form. Otherwise its integral would become negative when the orientation of a phase space volume changes, leading to unphysical negative energy.

Remark 3.20. The phase space light energy density ℓ is defined using the Liouville form $\varpi = dp \wedge dq$ that is invariant under a Hamiltonian flow. In many instances, it is hence sufficient to study the scalar component $\mathcal{L}(q,p)$, analogous to the treatment of densities in classical continuum mechanics in $\mathbb{R}^3$ where also an invariant volume form exists, cf. Theorem 2.23. However, care is required since functions and volume forms do behave differently under a change of variables, and the appropriate Jacobian terms have to be considered for the density ℓ.

We saw in Chapter 3.2.4 that the dynamics of light transport can be reduced from the slit cotangent bundle $T^*Q \setminus \{0\}$ to the cosphere bundle S^*Q, at least when only a fixed frequency is considered. Instead of the canonical coordinates (q,p) employed in the foregoing, it is hence useful to employ a parametrization of the fibers $T_q^*Q \setminus \{0\}$ in spherical coordinates, which for example allows an elementary description of the cosets $[p] \in S^*Q$ as radial sections, see again Fig. 3.7.

The standard change of variables for a density from Cartesian to spherical coordinates is given by[134]

$$dp = \frac{1}{J_{c \to s}}\, d\theta\, d\phi\, dr = \|r\|^2 \sin(\theta)\, d\theta\, d\phi\, dr = \|r\|^2\, d\bar{p}\, dr. \qquad (3.102)$$

---

[133]The compactness of ℓ can also be justified more intrinsically using the compactness of the electromagnetic field and standard results in semi-classical analysis; see Theorem 4.28 and the ensuing remark in (Evans and Zworski, *Semiclassical Analysis*).

[134]We will avoid writing ω for a point on the sphere and $d\omega$ for the associated surface element to avoid confusion with the angular frequency, and we will instead employ $\bar{p}$ and $d\bar{p}$. The symplectic 2-form is also denoted by ω but it should always be clear from the context which concept we refer to.

For us this takes the form

$$dp = \|p\|^2 \, d\bar{p} \, dr = \frac{n^2(q)}{c^2} \, 4\pi^2 \, \nu^2 \, d\bar{p} \, dr \tag{3.103}$$

and the change of variables is a diffeomorphism since we excluded the zero section from the fibers. Hence, the light energy density parametrized in spherical coordinates is

$$\ell = \mathcal{L}(q, \bar{p}, r) \, \frac{n^2(q)}{c^2} 4\pi^2 \nu^2 \, dq \, d\bar{p} \, dr. \tag{3.104}$$

Above, the radial direction is parametrized by an abstract length scale r. However, physically more natural is a parametrization by frequency ν. With Eq. 3.24, the change of variables from r to ν is

$$dr = \frac{1}{J_{r \to \nu}} d\nu = \frac{n(q)}{c} \, 2\pi \, d\nu. \tag{3.105}$$

For the light energy density parametrized by frequency we hence obtain

$$\ell = \mathcal{L}(q, \bar{p}, \nu) \, \frac{n^3(q)}{c^3} \, 8\pi^3 \nu^2 \, dq \, d\bar{p} \, d\nu. \tag{3.106}$$

Subsuming the spatially invariant terms into the coordinate function yields

$$\ell = n^3(q) \, \tilde{\mathcal{L}}(q, \bar{p}, \nu) \, dq \, d\bar{p} \, d\nu. \tag{3.107}$$

where we defined the pre-radiance function $\tilde{\mathcal{L}}$ as

$$\tilde{\mathcal{L}} = \mathcal{L}(q, \bar{p}, \nu) \, \frac{1}{c^3} \, 8\pi^3 \nu^2. \tag{3.108}$$

We will also introduce the pre-radiance density

$$\tilde{\ell} \equiv \tilde{\ell}_\nu = \frac{1}{n^3(q)} \ell = \tilde{\mathcal{L}}(q, \bar{p}, \nu) \, dq \, d\bar{p} \, d\nu \in \mathrm{Den}(T^*Q \backslash \{0\}) \tag{3.109}$$

which is the part of the light energy density parametrized in spherical coordinates that is independent of the refractive index. When the refractive index is unity, as it is in vacuum or in good approximation in air, then light energy density and pre-radiance coincide, and the above nomenclature is justified since the measurement of $\tilde{\ell}$ will yield what is classically known as radiance, as will be discussed in detail in the next section.[135]

Remark 3.21. A parametrization of the light energy density in spherical coordinates is natural, at least to a certain extent, when the dynamics are to be reduced from the slit cotangent bundle $T^*Q \backslash \{0\}$ to the cosphere bundle

---

[135]A cubic dependence on the refractive index might appear surprising at first sight but it was pointed out before in the literature, see for example (Liebes, "Brightness–On the Ray Invariance of B/n^2", p. 933).

S^*Q since this enables to easily choose a radial sections in $T_q^*Q\setminus\{0\}$ which fixes a representative for each coset $[p]$. However, the cubic dependence on the refractive index make pre-radiance $\tilde{\ell}$ less suited when time evolution is considered. This suggests a definition of pre-radiance $\tilde{\ell}$ subsuming the $n^3(q)$ factor. It would not allow to recover a well known quadratic dependence of radiance on the refractive index, cf. Chapter 3.2.6.3.

Remark 3.22. A parametrization of the quantities representing the transported energy over the sphere or cosphere bundle is also employed in the classical literature. It is interesting to observe that the contact structure introduced in the last section allows to mathematically justify this parametrization, despite the fact that the system is also naturally Hamiltonian.[136]

As an alternative to Eq. 3.107, one could also introduce pre-radiance parametrized with respect to wavelength $\lambda = v(q)\,\nu = c/n(q)\,\nu$. The Jacobian is then

$$J_{\nu\to\lambda} = -\frac{c}{n(q)}\frac{1}{\nu^2} = -\frac{c}{n(q)}\frac{n^2(q)}{c^2}\lambda^2 = -\frac{n(q)}{c}\lambda^2 \tag{3.110}$$

so that from Eq. 3.106 one obtains

$$\ell = \mathcal{L}(q,\bar{p},\nu)\,\frac{n^3(q)}{c^3}\,8\pi^3\nu^2\;dq\,d\bar{p}\,\frac{1}{\lambda^2}\frac{c}{n(q)}d\lambda \tag{3.111a}$$

$$= \mathcal{L}(q,\bar{p},\nu)\,\frac{n^2(q)}{c^2}\,8\pi^3\frac{\nu^2}{\lambda^2}\;dq\,d\bar{p}\,d\lambda \tag{3.111b}$$

where we also used that the phase space energy density is a 1-density and not a volume form, and hence the absolute value of the Jacobian has to be employed. Pre-radiance density $\tilde{\ell}_\lambda \in \mathrm{Den}(T^*Q\setminus\{0\})$ parametrized with respect to wavelength is hence given by

$$\tilde{\ell}_\lambda(q,\bar{p},\nu) = \frac{1}{n^2(q)}\ell(q,p) = \tilde{\mathcal{L}}_\lambda(q,\bar{p},\lambda)\,dq\,d\bar{p}\,d\lambda. \tag{3.112}$$

At first sight, it might seems surprising that $\tilde{\ell}_\nu$ has a cubic dependence on the refractive index while $\tilde{\ell}_\lambda$ has a quadratic dependence. However, since frequency is conserved during transport and

$$d\lambda = \frac{c}{n(q)}\nu^2 d\nu \tag{3.113}$$

we can think of $d\lambda$ as being locally scaled by the refractive index, which explains the discrepancy between $\tilde{\ell}_\nu$ and $\tilde{\ell}_\lambda$. This also shows that a parametrization by

---

[136]Gershun gives the following rationale for a parametrization in spherical coordinates: *"Since the orientation of the surface element is completely and uniquely defined by the direction of the surface normal, the illumination may be regarded as a function of two factors: position in space, and direction."*, (Gershun, "The Light Field", p. 59). It has to be observed that for Gershun the directional dependence arises through surfaces, and hence through measurements, which so far do not play a role for us.

frequency and not wavelength is more naturally when transport is considered since then no local scaling takes place.

Remark 3.23. From the literature, it is not apparent if a parametrization of the phase space energy density in spherical coordinates should employ frequency, angular frequency, or wavelength. There, (pre-)radiance is usually not considered as a spectral quantity, and the various fields employing radiometry, cf. Chapter 3.1, often employ different interpretations. Sometimes the concept of spectral radiance can be found in the literature but the lack of a formal connection to radiance prevents further insight.[137] Physically, the parametrization is irrelevant in that the measured energy flux is independent of the chosen scaling in the radial fibers.

Remark 3.24. With the qualifications in the previous remark in mind, the pre-radiance density $\tilde{\ell}$ can be considered as what is in the classical literature sometimes denoted as the plenoptic function or light field.[138]

Analogous to Eq. 3.51a for the light energy density ℓ, the transport of pre-radiance $\tilde{\ell}$ can be described using the pullback

$$\eta_t^* \left(n^3 \tilde{\ell}_t \right) = n^3 \tilde{\ell}_0 \tag{3.114}$$

along the flow η_t generated by the Hamiltonian vector field X_H. The pullback provides a generalization of the ray casting or boundary distance function,[139] and it is recovered when one chooses t as the maximal value for which the flow η_t is defined. In contrast to the classical formulation, however, the pullback is available in media with varying refractive index and it is mathematically well founded, which enables to combine it with other operations such as differentiation.

Similar to the light energy density, a transport equation for pre-radiance $\tilde{\ell}$ can be obtained with the coordinate expression of the Poisson bracket,

$$\frac{\partial n^3 \tilde{\ell}}{\partial t} = -\left\{ n^3 \tilde{\ell}, H \right\} = -\frac{\partial\, n^3 \tilde{\ell}}{\partial q} \cdot \frac{\partial H}{\partial p} + \frac{\partial\, n^3 \tilde{\ell}}{\partial p} \cdot \frac{\partial H}{\partial q}. \tag{3.115}$$

For the first term we have by the Leibniz rule and with the explicit expression for the Hamiltonian vector field in Eq. 3.40 that

$$\frac{\partial n^3 \tilde{\ell}}{\partial q} \cdot \frac{\partial H}{\partial p} = \left(3\, n^2 \tilde{\ell} \frac{\partial n}{\partial q} + n^3 \frac{\partial \tilde{\ell}}{\partial q} \right) \cdot \frac{c}{n} \frac{p}{\|p\|} = c\, n \left(3\, \tilde{\ell} \frac{\partial n}{\partial q} + n \frac{\partial \tilde{\ell}}{\partial q} \right) \cdot \frac{p}{\|p\|}. \tag{3.116}$$

---

[137]Nicodemus, "Radiance"; Goodman, "General Principles of Geometrical Optics".

[138]The concept of the light field was introduced in (Gershun, "The Light Field"), see the recent article by Levoy ("Light Fields and Computational Imaging") for a contemporary treatment. The plenoptic function was introduced by Adelson and Bergen ("The Plenoptic Function and the Elements of Early Vision").

[139](Veach, "Robust Monte Carlo Methods for Light Transport Simulation", p. 90 and p. 110) and (Arvo, "Analytic Methods for Simulated Light Transport", p. 136).

354

Using the chain rule and the usual expression for the gradient in spherical coordinates (r, θ, ϕ), we obtain for the partial derivative with respect to momentum in the second term of Eq. 3.115 that

$$\frac{\partial n^3 \tilde{\ell}}{\partial p} = n^3 \left(\frac{\partial \tilde{\ell}}{\partial \nu} \frac{\partial \nu}{\partial p} + \frac{1}{\nu} \frac{\partial \tilde{\ell}}{\partial \theta} \frac{\partial \theta}{\partial p} + \frac{1}{\nu \sin \theta} \frac{\partial \tilde{\ell}}{\partial \phi} \frac{\partial \phi}{\partial p} \right). \tag{3.117}$$

The first term in Eq. 3.117 vanishes when inserted in Eq. 3.115, since frequency is conserved during transport, and hence

$$\frac{\partial n^3 \tilde{\ell}}{\partial p} = \frac{n^3}{\nu} \left(\frac{\partial \tilde{\ell}}{\partial \theta} \frac{\partial \theta}{\partial p} + \frac{1}{\sin \theta} \frac{\partial \tilde{\ell}}{\partial \phi} \frac{\partial \phi}{\partial p} \right) = \frac{n^3}{\|p\|} \frac{\partial}{\partial p_t} \tilde{\ell} \tag{3.118}$$

where $r = \|p\|$ since we have to employ spherical coordinates with respect to p for the partial derivatives $\partial/\partial p$, cancelling the additional scaling that leads to parametrization by frequency. In Eq. 3.118 we also introduced the operator $\partial/\partial p_t$ for the tangential part of the derivative. With the explicit expression for $\partial H/\partial q$ from the Hamiltonian vector field in Eq. 3.40, the second term of Eq. 3.115 hence becomes

$$\frac{\partial n^3 \tilde{\ell}}{\partial p} \cdot \frac{\partial H}{\partial q} = \frac{n^3}{\|p\|} \frac{\partial}{\partial p_t} \tilde{\ell} \cdot \frac{c}{n^2} \|p\| \frac{\partial n}{\partial q} = c\, n \frac{\partial}{\partial p_t} \tilde{\ell} \cdot \frac{\partial n}{\partial q}. \tag{3.119}$$

For the Poisson bracket in Eq. 3.115 we thus obtain

$$-\left\{ n^3 \tilde{\ell}, H \right\} = -c\, n \left(3 \tilde{\ell} \frac{\partial n}{\partial q} + n \frac{\partial \tilde{\ell}}{\partial q} \right) \cdot \frac{p}{\|p\|} + c\, n \frac{\partial}{\partial p_t} \tilde{\ell} \cdot \frac{\partial n}{\partial q} \tag{3.120}$$

and using ∇-notation and after collecting terms the transport equation for pre-radiance in a medium with varying refractive index is given by[140]

$$\frac{n}{c} \frac{\partial \tilde{\ell}}{\partial t} + \frac{p}{\|p\|} \cdot \nabla_q \tilde{\ell} - \frac{1}{n} \nabla_q n \cdot \nabla_{p_t} \tilde{\ell} + \frac{3 \tilde{\ell}}{n} \nabla_q n \cdot \frac{p}{\|p\|} = 0. \tag{3.121}$$

Compared to Eq. 3.53c which describes the transport of the phase space energy density ℓ, Eq. 3.121 contains an additional term that arises from the explicit dependence on the refractive index. In a homogeneous medium and for a system in steady state where the Eulerian rate of change $\partial \ell/\partial t$ vanishes, Eq. 3.121 becomes

$$\frac{p}{\|p\|} \cdot \nabla_q \tilde{\ell} = \bar{s} \cdot \nabla_q \tilde{\ell} = 0 \tag{3.122}$$

---

[140]Bal ("Radiative transfer equations with varying refractive index: a mathematical perspective") obtains a different expression for the transport equation for radiance but he also employs a different definition of (pre-)radiance. We will comment on the difference in detail in Chapter 3.2.6.3.
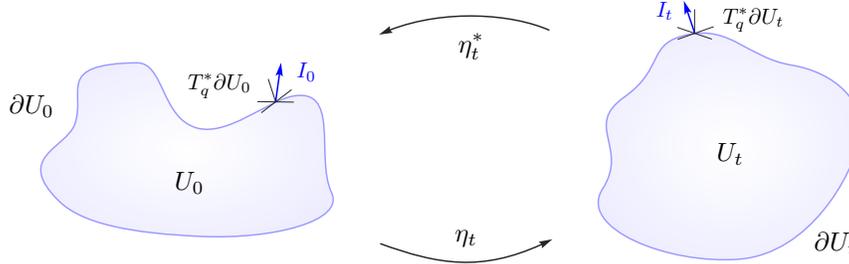
Figure 3.9: Geometry of the transport theorem used to derive the measurement equation for light transport.

which shows that pre-radiance is constant along a ray in such environments.[141]

The above derivation and the constancy of the light energy density ℓ but not of pre-radiance density $\tilde{\ell}$ shows that in media with spatially varying refractive index the phase space energy density is the quantity most naturally studied and considered.

3.2.6 Measurement of Phase Space Energy Density

In the following, we will study the measurement of phase space energy density and pre-radiance. Similar to the derivation of the continuity equation in fluid dynamics, cf. Remark 2.116, we will consider the transport theorem for a volume in phase space, and then employ Stokes' theorem to relate the transport in the volume to the flux through the surface. The derivation will also illuminate various concepts from the classical literature, such as the cosine term and radiance, cf. Chapter 3.2.6.3. Our derivation will emphasize intuition over technicalities, and some of the details are therefore deferred to Chapter 3.2.6.4.

3.2.6.1 The Transport Theorem

For light energy density ℓ, the transport theorem in Theorem 2.23 is given by

$$\frac{d}{dt} \int_{\eta_t(W)} \ell = \int_{\eta_t(W)} \left(\frac{\partial \ell}{\partial t} + \pounds_{X_H} \ell \right) \tag{3.123}$$

where $W = U \times \mathbb{R}^3 = T^*U \subset T^*Q$ is an arbitrary paracompact volume in phase space, cf. Remark 2.119 and also Fig. 3.9. In Eq. 3.123, the flow η_t is generated by the Hamiltonian vector field X_H in Eq. 3.40, and it is a local diffeomorphism as discussed in Remark 3.2. With the conservation of the light energy density along trajectories in phase space, the left hand side of Eq. 3.123 vanishes and

---

[141]In the computer graphics literature, the analogous expression for radiance can for example be found in (Arvo, "Transfer Equations in Global Illumination", Eq. 39).

writing $W_t = \eta_t(W)$ we therefore obtain

$$\int_{W_t} \frac{\partial \ell}{\partial t} = \int_{W_t} \pounds_{X_H} \ell \tag{3.124}$$

where the left hand side now represents the total Eulerian rate of change for the phase space volume W_t at the time t. By Cartan's formula, the transport of the volume form ℓ is

$$\pounds_{X_H} \ell = \mathrm{d} \mathrm{i}_{X_H} \ell + \mathrm{i}_{X_H} \mathrm{d} \ell = \mathrm{d} \mathrm{i}_{X_H} \ell. \tag{3.125}$$

and the transport theorem thus takes the form

$$\int_{W_t} \frac{\partial \ell}{\partial t} = \int_{W_t} \mathrm{d} \mathrm{i}_{X_H} \ell. \tag{3.126a}$$

With Stokes' theorem we obtain

$$\int_{W_t} \frac{\partial \ell}{\partial t} = \int_{\partial W_t} \mathrm{i}_{X_H} \ell. \tag{3.126b}$$

and the six dimensional integral over W_t is reduced to the integral of the five dimensional volume form $\mathrm{i}_{X_H} \ell \in \Omega^5(\partial W_t)$ over the boundary ∂W_t. Intuitively, Eq. 3.126b states that the change in the phase space light energy density ℓ in W_t is given by the flow $\mathrm{i}_{X_H} \ell$ of ℓ through the boundary ∂W_t generated by the Hamiltonian vector field X_H, although some care is required with this interpretation since W_t is a paracompact volume in phase space; see also the ensuing remark.

Remark 3.25. Intuition for the interior product $\mathrm{i}_{X_H} \ell$ can be obtained when the derivation of Gauss theorem is considered. There, the interior product $\mathrm{i}_X \, dx^3$ for the canonical volume form dx^3 on Euclidean space arises. Using the metric on $\mathbb{R}^3$, $\mathrm{i}_X \, dx^3$ can be written as $X \cdot \vec{n} \, dA$, which is the integrand when the vector field $X \in \mathfrak{X}(R^3)$ is integrated over a boundary surface with normal $\vec{n}$. Alternatively, insight into $\mathrm{i}_{X_H} \ell$ can also be obtained by considering the phase space density as $\ell = \mathcal{L} \mu$ where μ is the Euclidean volume form on $T^*Q \cong \mathbb{R}^3 \times \mathbb{R}^3$. Then

$$\mathrm{i}_{X_H} \ell = \mathrm{i}_{X_H} \mathcal{L} \mu = \mathrm{i}_{X_H}(\mathcal{L} \wedge \mu) = \mathrm{i}_{X_H} \mathcal{L} \wedge \mu + \mathcal{L} \wedge \mathrm{i}_{X_H} \mu = \mathcal{L} \, \mathrm{i}_{X_H} \mu$$

and with the metric on $\mathbb{R}^3 \times \mathbb{R}^3$ one obtains, analogous to Remark 2.107, that

$$\mathcal{L} \, \mathrm{i}_{X_H} \mu \cong \mathcal{L}(\star \mathrm{i}_{X_H} \mu)^\sharp = \mathcal{L} X_H.$$

We can hence think of $\mathrm{i}_{X_H} \ell$ as the Hamiltonian vector field scaled or weighted by the phase space density. For a medium with homogeneous refractive index where $X_H = (\dot{q}, \dot{p}) \propto (p/\|p\|, 0)$ the interior product $\mathrm{i}_{X_H} \ell$ can be identified with the density $\ell_t(q, p) \, p/\|p\|$ over unit norm vector fields on configuration space, or equivalently with a density over the sphere bundle SQ. The interior product thus also corresponds to the intuitive picture one has in mind for straight rays whose magnitude is given by the transported energy.

The boundary of the phase space volume $W_t = T^*U_t = U_t \times \mathbb{R}^3$ in Eq. 3.126 is given by $\partial W_t = \partial U_t \times \mathbb{R}^3$, cf. Remark 2.50, and hence the equation can be written as

$$\int_{W_t} \frac{\partial \ell}{\partial t} = \int_{T^*\partial U_t} i_{X_H}\ell \tag{3.127a}$$

$$= \int_{\partial U_t} \int_{T_q^*\partial U_t} i_{X_H}\ell \tag{3.127b}$$

$$= \int_{\partial U_t} \bar{\bar{\ell}} \tag{3.127c}$$

where $T_q^*\partial U$ are the fibers over the boundary ∂U in configuration space, cf. again Fig. 3.9. As shown in Chapter 3.2.6.4, the volume form $\bar{\bar{\ell}} \in \Omega^2(\partial U_t) \cong \mathrm{Den}(\partial U)$ over the boundary surface ∂U_t is given by

$$\bar{\bar{\ell}} = \int_{T_q^*\partial U_t} i_{X_H}\ell = \quad J^1(u)\, n_1(u)\, du^2 \wedge du^3$$
$$+ J^2(u)\, n_2(u)\, du^1 \wedge du^3 \tag{3.128}$$
$$+ J^3(u)\, n_3(u)\, du^1 \wedge du^2$$

where $u = (u^1, u^2)$ is a local parametrization for the surface ∂U, $\vec{n} = (n_1, n_2, n_3)$ is the local surface normal, and we assume, without loss of generality, that $\vec{n}$ is of unit length. The components $J^j(u)$ in Eq. 3.128 are

$$J^j(u) = \int_{T_q^*\partial U_t} \mathcal{L}(u, p)\, X_q^j(u, p)\, dp \tag{3.129}$$

and the $X_q^j = \dot{q}^j$ are the configuration space components of the Hamiltonian vector field X_H. Using the explicit expression for X_H from Eq. 3.40 and combining Eq. 3.128 and Eq. 3.129 we obtain

$$\bar{\bar{\ell}} = \left(\int_{T_q^*\partial U_t} \mathcal{L}(u, p)\, \frac{c}{n(q)} \frac{p}{\|p\|}\, dp \right) \cdot \vec{n}\, dA \tag{3.130a}$$

$$= \left(\frac{c}{n(q)} \int_{T_q^*\partial U_t} \mathcal{L}(u, p)\, (\bar{p} \cdot \vec{n})\, dp \right) dA \tag{3.130b}$$

where $\bar{p} = p/\|p\|$. The factor of $c/n(q)$ in Eq. 3.130b is the speed of light in the medium and it can be considered as the flux rate through the boundary ∂U_t. Using pre-radiance instead of the phase space energy density we obtain for Eq. 3.128

$$\bar{\bar{\ell}} = \int_{T_q^*\partial U_t} i_{X_H}(n^3\tilde{\ell}) = c\,n^2(\quad \tilde{J}^1(u)\, n_1(u)\, du^2 \wedge du^3$$
$$+ \tilde{J}^2(u)\, n_2(u)\, du^1 \wedge du^3 \tag{3.131}$$

$$+ \tilde{J}_3(u)\, n_3(u)\, du^1 \wedge du^2)$$

with the components $\tilde{J}^j$ given by

$$\tilde{J}^j = \int_{S_q^* U_t} \int_{\mathbb{R}^+} \frac{p^j}{\|p\|}\, \tilde{L}(u, \bar{p}, \nu)\, d\bar{p}\, d\nu \qquad (3.132)$$

where we immediately used the explicit expression for X_q^j from the Hamiltonian vector field X_H. Analogous to Eq. 3.130, with pre-radiance the fiber integral $\bar{\bar{\ell}}$ can hence also be written as

$$\bar{\bar{\ell}} = c\, n^2(q) \left(\int_{S^2} \int_{\mathbb{R}^+} \tilde{L}(u, \bar{p}, \nu)(\bar{p} \cdot \vec{n})\, d\bar{p}\, d\nu \right) dA \qquad (3.133)$$

where we also identified the cosphere bundle with the sphere bundle using the canonical metric on Euclidean space.

3.2.6.2 Measurement of Radiant Energy

Motivated by the above derivation, we define the measurement of phase space energy density over a non-closed, opaque surface $\mathcal{M} \subset \mathbb{R}^3$ in time $[t_1, t_2]$ as

$$E = \int_{t_1}^{t_2} \int_{T^-\mathcal{M}} i_{X_H} \ell \qquad (3.134a)$$

$$= \int_{t_1}^{t_2} \int_{\mathcal{M}} \bar{\bar{\ell}} \qquad (3.134b)$$

$$= \int_{t_1}^{t_2} \int_{\mathcal{M}} \frac{c}{n(q)} \int_{T_q^-\mathcal{M}} \mathcal{L}(u, p)\,(\bar{p} \cdot \vec{n})\, dp\, dA\, dt \qquad (3.134c)$$

where $T^-\mathcal{M}$ denotes the half space of $T^*\mathcal{M}$ where $(\bar{p} \cdot \vec{n})$ is negative when considered fiber-by-fiber, cf. Def. 2.84, and $T^*\mathcal{M}$ is the restriction $T^*\mathcal{M} = T^*Q|_{\mathcal{M}}$ of T^*Q to the surface $\mathcal{M}$, and it is sufficient to consider $T^-\mathcal{M}$ since it is the subset of phase space where the phase space light energy density is nontrivial. Analogously, the measurement of radiant energy using pre-radiance is

$$E = \int_{t_1}^{t_2} \int_{T^-\mathcal{M}} i_{X_H}(n^3 \ell) \qquad (3.135a)$$

$$= \int_{t_1}^{t_2} \int_{\mathcal{M}} c\, n^2(q) \int_{S_q^-\mathcal{M}} \int_{\mathbb{R}^+} \tilde{L}(u, \bar{p}, \nu)(\bar{p} \cdot \vec{n})\, d\bar{p}\, d\nu\, dA\, dt \qquad (3.135b)$$

where $S_q^-\mathcal{M}$ denotes the negative hemisphere as defined by the surface normal of $\mathcal{M}$, and the equation is easily modified for monochromatic radiation.

We will refer to Eq. 3.134 and Eq. 3.135b as the **measurement equations** for light transport.[142] It follows from the foregoing derivation and the left hand side of Eq. 3.126, representing the Eulerian rate of change $\partial \ell / \partial t$ of energy in the phase space volume W_t, that measurements yield the energy flowing through a surface in time $[t_1, t_2]$. As mentioned before, the factor of $v(q) = c/n(q)$ in Eq. 3.134c thus determines the flux rate through the surface.

3.2.6.3   Radiometry Revisited

The derivation of the measurement equation from the transport theorem presented in the foregoing enables to relate the phase space light energy density and pre-radiance to concepts from classical radiometry, such as radiance and vector irradiance. We will develop these connections in this section.

Radiance   In classical radiometry, the central quantity of interest is radiance which is often described as "radiant flux or power per unit solid-angle-in-the-direction-of-a-ray per unit projected-area-perpendicular-to-the-ray".[143] As discussed in Chapter 3.1, the concept was introduced as the infinitesimal 'splendor' by Lambert in his *Photometria*, which was founded in and developed from the measurement experiments he performed. Hence, one expects that radiance can be recovered from our formulation when the infinitesimal description of measurements is considered.

Using pre-radiance, measurements are infinitesimally described by the integrand of Eq. 3.135b. By linearity, the integrand can be written as

$$\tilde{L}(u, \bar{p}, \nu)(\bar{p} \cdot \vec{n}) \, d\bar{p} \, d\nu \, dA = \vec{n} \cdot \left(\tilde{L}(u, \bar{p}, \nu) \, d\bar{p} \, d\nu \, \bar{p} \, dA \right) \tag{3.136}$$

and we separated the surface-independent quantities from the surface-dependent normal. To obtain a more intrinsic description, we define the 2-form

$$dA_\perp = \bar{p}_1 du^2 \wedge du^3 + \bar{p}_2 du^1 \wedge du^3 + \bar{p}_2 du^1 \wedge du^2 \in \Omega^2(Q) \tag{3.137}$$

where the $\bar{p}_i$ are as usual the components of the normalized momentum $\bar{p}$. The surface independent part of Eq. 3.136 can then be written as

$$\tilde{L}(u, \bar{p}, \nu) \, d\bar{p} \, d\nu \, dA_\perp \tag{3.138}$$

and the original equation is recovered when the pullback of the 2-form in Eq. 3.138 onto a surface $\mathcal{M}$ with normal $\vec{n}$ is considered, cf. Rem. 2.114. The 2-form $dA_\perp$ can be interpreted as flux through a surface orthogonal to the normalized momentum $\bar{p}$ since the cosine term arising from the pullback then becomes unity. With this interpretation of $dA_\perp$ and since Eq. 3.138

---

[142]The definition of the measurement equation is consistent with those in kinetic theory, see for example (Cercignani, *The Boltzmann Equation and its Applications*, p. 57), and computer graphics, for example (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*, p. 760) or (Dutré, Bala, and Bekaert, *Advanced Global Illumination*, p. 45).

[143]Nicodemus, "Radiance".

represents infinitesimal measurements when pulled back onto a surface, we identify **classical radiance** L with[144]

$$L(u, \nu, \bar{p}) \equiv \tilde{L}(u, \bar{p}, \nu) \, d\bar{p} \, d\nu \, dA_\perp \tag{3.139}$$

and it indeed represents the "radiant flux or power per unit solid-angle-in-the-direction-of-a-ray per unit projected-area-perpendicular-to-the-ray".[145]

In the previous section, we introduced the notion of pre-radiance $\tilde{\ell}$, which to our knowledge did not appear before in the literature. With Eq. 3.139, our definition and nomenclature was hence motivated by the connection to classical radiance that arises when measurements are considered.[146] It should be noted that while it is possible to relate radiance at two different points in space or along a ray, and to denote this as transport, what is intrinsically transported is a density over phase space, the phase space energy density ℓ or pre-radiance $\tilde{\ell}$, and not radiance.[147]

Remark 3.26. The importance of radiance for conceptual considerations and for computations has been emphasized throughout the literature.[148] However, to our knowledge no physically and mathematically satisfying definition has been proposed so far. In particular, most recent literature[149] has failed to notice that radiance is associated with measurements, although this seems apparent from Lambert's work where it was introduced, cf. Chapter 3.1 and Fig. 3.1, and this was also well known in the past.[150] The confusion is well exemplified by a recent derivation of the transport equation by Bal where radiance is defined as[151]

$$\int_{S^2} \int_{\mathbb{R}^+} \frac{n(q)}{c} L(x, \omega, \nu) \, d\omega \, d\nu = \int_{T_q^* Q} \mathcal{L}(q, p) \, dp. \tag{3.140}$$

Bal provides a lengthy justification for the explicit factor of $n(q)/c$ on the left hand side of Eq. 3.140: In the literature radiance has units of $J/m^2 \, sr \, Hz$ where

---

[144]See for example (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*, p. 285) or (Dutré, Bala, and Bekaert, *Advanced Global Illumination*, p. 20).

[145]Nicodemus, "Radiance".

[146]One could ask why we do not define radiance as the flux through a chosen surface. However, this would not allow to recover the cosine term which with our definition arises naturally through the pullback.

[147]In his *Photometria*, Lambert avoided a discussion of transport processes and considered his work as a phenomenological explanation of the observed measurements.

[148]See for example (Pomraning, *The Equations of Radiation Hydrodynamics*; Wolf, "Coherence and Radiometry"; Arvo, "Analytic Methods for Simulated Light Transport"; Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*).

[149]See for example (Wolf, "Coherence and Radiometry"; Bal, "Radiative transfer equations with varying refractive index: a mathematical perspective"; Zhang and Levoy, "Wigner Distributions and How They Relate to the Light Field").

[150]See for example (Liebes, "Brightness–On the Ray Invariance of B/n^2") and Gershun makes it also clear when he says: "Since the orientation of the surface element is completely and uniquely defined by the direction of the surface normal, the illumination may be regarded as a function of two factors: position in space, and direction.", (Gershun, "The Light Field", p. 59), and hence a surface is assumed.

[151](Bal, "Radiative transfer equations with varying refractive index: a mathematical perspective", Eq. 13), using our nomenclature.

J represents Joule, m is meter, sr stands for steradians, the artificial unit associated with the solid angle, and $Hz = 1/s$ is the unit of frequency. Using classical conventions, the units for the phase space energy density are however $J/m^3 \, m^3$ or, according to Bal, $J/m^3 \, sr$ with a parametrization of the fiber in spherical coordinates.[152] To resolve the apparent mismatch between the cubic m^3 and quadratic m^2 dependence on space, Bal introduces the factor $n(q)/c$ in Eq. 3.140 that, since the refractive index $n(q)$ is unit-less, has units of s/m, and formally makes the units of radiance as defined in Eq. 3.140 agree with those found in the classical literature. However, as discussed above, the quadratic m^2 space dependence of radiance is justified because it represents a measurement through a surface while the phase space energy density has a cubic dependence on space since it is a density over the cotangent bundle $T^*Q \cong \mathbb{R}^3 \times Q \subset \mathbb{R}^3$.

Remark 3.27. In the literature,[153] radiance or radiant intensity are sometimes identified with the magnitude of the Poynting vector for a plane wave which is given by[154]

$$\|\vec{S}(q)\| = \left\| \frac{c}{n(q)} \mathcal{E}(q,t) \, \bar{s} \right\| = \frac{c}{n(q)} \mathcal{E}(q,t). \tag{3.141}$$

Using the identification of the Poynting vector with the first moment of the phase space energy density in Eq. 3.47b, we have for a single direction that[155]

$$\|\vec{S}(q)\| = \left\| \frac{c}{n(q)} \frac{p}{\|p\|} \mathcal{L}(q,p) \right\| = \left\| \frac{c}{n(q)} \mathcal{L}(q,p) \, \bar{s} \right\| = \frac{c}{n(q)} \mathcal{L}(q,p). \tag{3.142}$$

Hence, in vacuum, and to good approximation in air, the identification of the phase space energy density function $\mathcal{L}$ with $\|\vec{S}\|$ is valid.

The Cosine Term The measurement equation in Eq. 3.134 contains the factor $(\bar{p} \cdot \vec{n})$ which in the literature is known as cosine or foreshortening term, cf. also Chapter 1.1.1. Although it is prevalent in light transport theory, it has so far only be justified using heuristic, pre-Descartian arguments.[156] Using our ansatz, it is an immediate consequence of the transport theorem, a central pillar of continuum mechanics, and it arises canonically from the pullback of a 2-form onto a surface, cf. Remark 2.114.

---

[152]More correctly, the units should be $J/m^3 \, sr \, Hz$ since the radial component is associated with frequency and otherwise also one dimension would be "lost".

[153]See for example (Born and Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Chapter 3.1.4) and (Marathay and McCalmont, "Diffraction").

[154]Born and Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Chapter 1.4.1.

[155]More formally the Dirac delta distribution should be employed as a test function to study the behaviour for a single direction.

[156]See for example (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*, Chapter 5.4) or (Dutré, Bala, and Bekaert, *Advanced Global Illumination*, Chapter 2.3).

362

Vector Irradiance and the Poynting Vector   In Eq. 3.129 the vector components $J^j(u)$ were defined but we have not discussed their physical significance. With the expression for the Hamiltonian vector field X_H from Eq. 3.40, the definition becomes

$$J(u) = \int_{T_q^* \partial U} \mathcal{L}(u,p) \, \frac{c}{n(q)} \frac{p}{\|p\|} \, dp = \frac{c}{n(q)} \int_{T_q^* \partial U} \frac{p}{\|p\|} \, \mathcal{L}(u,p) \, dp \qquad (3.143)$$

which is the first moment of ℓ and the expression for the Poynting vector in terms of the phase space light energy density in the literature, cf. Remark 3.6. Additionally, $J(u)$ can also be identified with **classical vector irradiance** $\vec{I}$ since Eq. 3.128 is then equivalent to

$$\bar{\bar{f}} = I(q) = \vec{I} \cdot \vec{n} \, dA \qquad (3.144)$$

which can be considered as the classical definition of the concept.[157] The volume form $I(q) \in \mathrm{Den}(\mathcal{M})$ on the surface $\mathcal{M}$ in Eq. 3.144 is classically known as **irradiance**, and it can also be defined directly through the fiber integral in the measurement equation as

$$I(q) = \int_{\mathbb{H}^+} \int_{\mathbb{R}^+} \tilde{L}(u,\bar{p},\nu)(\bar{p} \cdot \vec{n}) \, d\bar{p} \, d\nu \; \in \mathrm{Den}(\mathcal{M}). \qquad (3.145)$$

Next to the equivalence of Poynting's theorem and the transport equation for the phase space energy density discussed in Remark 3.6, we hence also have the equivalence of the Poynting vector with vector irradiance defined using the phase space energy density. $\tilde{J}(u)$ in Eq. 3.132, which was the analogue of $J(u)$ with respect to pre-radiance, could also be identified with vector irradiance and this would arguably be more consistent with the classical literature. However, Eq. 3.144 would then only hold when the refractive index is unity and hence we consider it as more appropriate to define vector irradiance based on the phase space energy density.

Basic Radiance   Eq. 3.135b shows the well known quadratic dependence of radiance on the refractive index.[158] However, some care is required on how this dependence is to be interpreted.

Let t_1 and t_2 be two arbitrary times with $t_2 > t_1$. For a system in steady state where the Eulerian rate of change $\partial \ell / \partial t$ vanishes we have from Eq. 3.124 that

$$\int_{W_{t_1}} \pounds_{X_H} \ell = \int_{W_{t_2}} \pounds_{X_H} \ell \qquad (3.146)$$

---

[157]Vector irradiance was introduced by Gershun ("The Light Field") as 'light vector' and has been used in computer graphics for example for irradiance caching (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*, Chapter 15.5), following work by Ward and Heckbert ("Irradiance Gradients") and Tabellion and Lamorlette ("An approximate global illumination system for computer generated films"), by Arvo ("The Irradiance Jacobian for Partially Occluded Polyhedral Sources"; "Analytic Methods for Simulated Light Transport") and by Stam ("Multiple scattering as a diffusion process").

[158]See for example (Liebes, "Brightness–On the Ray Invariance of B/n^2"; Goodman, "General Principles of Geometrical Optics").

and using Cartan's formula and Stokes' theorem yields

$$\int_{\partial W_{t_1}} i_{X_H}\ell = \int_{\partial W_{t_2}} i_{X_H}\ell. \tag{3.147}$$

With the same derivation that led to Eq. 3.127 and Eq. 3.133 we obtain infinitesimally for the 2-forms over the boundaries

$$c\,n^2(q_1)\left(\int_{S^2}\int_{\mathbb{R}^+}\tilde{L}(u_1,\bar{p}_1,\nu)(\bar{p}_1\cdot\vec{n}_1)\,d\bar{p}\,d\nu\right)dA(q_1)$$

$$= c\,n^2(q_2)\left(\int_{S^2}\int_{\mathbb{R}^+}\tilde{L}(u_2,\bar{p}_2,\nu)(\bar{p}_2\cdot\vec{n}_2)\,d\bar{p}\,d\nu\right)dA(q_2). \tag{3.148}$$

and by considering the 2-forms in Q whose pullback yields the above equation when integrated over the fibers we have

$$c\,n^2(q_1)\tilde{L}(u_1,\bar{p}_1,\nu)\,d\bar{p}\,d\nu\,dA_\perp(q_1)$$

$$= c\,n^2(q_2)\tilde{L}(u_2,\bar{p}_2,\nu)\,d\bar{p}\,d\nu\,dA_\perp(q_2) \tag{3.149}$$

where $dA_\perp \in \Omega^2(Q)$ is as defined in Eq. 3.137. A comparison with the definition of radiance in Eq. 3.139 now immediately shows that

$$n^2(q_1)L(q_1,\bar{p}_1,\nu) = n^2(q_2)L(q_2,\bar{p}_2,\nu) \tag{3.150}$$

which is the quadratic dependence of radiance on the refractive index. The scaled radiance $n^2(q)L(q,\bar{p},\nu)$ is in the literature sometimes denoted as basic radiance. Previous derivations[159] were often obtained in the context of black body radiation but do not necessarily generalize to arbitrary settings.

3.2.6.4 Detailed Derivation

In the derivation in Chapter 3.2.6.1 we omitted some technical details and these will be provided in the following. Returning to Eq. 3.127, the equation is more explicitly written as

$$0 = \int_{\partial U}\int_{T_q^*\partial U} i^*\left(i_{X_H}\ell\right) \tag{3.151}$$

where $i : \partial W \to W : \partial U \times \mathbb{R}^3 \to \mathbb{R}^3 \times \mathbb{R}^3 \cong T^*U$ is the inclusion map from the boundary ∂W into W, which is required by Stokes' theorem in Theorem 2.24. Using the Leibniz rule for the interior product $i_X(\alpha \wedge \beta) = i_X\alpha \wedge \beta \pm \alpha \wedge i_X\beta$ term by term for the density

$$\ell = \mathcal{L}(q,p)dq \wedge dp = \mathcal{L}(q,p)\,dq^1 \wedge dq^2 \wedge dq^3 \wedge dp^1 \wedge dp^2 \wedge dp^3 \tag{3.152}$$

---

[159]To our knowledge the earliest derivation of the quadratic dependence of radiance on the refractive index is due to Clausius ("Ueber die Concentration von Wärme- und Lichtstrahlen und die Gränzen ihrer Wirkung") who called it an "eigenthümlichen, theoretisch interessanten Schlusse" (a peculiar, theoretically interesting result).

the interior product $i_{X_H}\ell$ of the light energy density ℓ with the Hamiltonian vector field

$$X_H(q,p) = (X_q^1, X_q^2, X_q^3, X_p^1, X_p^2, X_p^3) \qquad (3.153)$$

becomes

$$
\begin{aligned}
i_{X_H}\ell = \mathcal{L}(q,p)(\quad & X_q^1(q,p)\, dq^2 \wedge dq^3 \wedge dp^1 \wedge dp^2 \wedge dp^3 \\
+ & X_q^2(q,p)\, dq^1 \wedge dq^3 \wedge dp^1 \wedge dp^2 \wedge dp^3 \\
+ & X_q^3(q,p)\, dq^1 \wedge dq^2 \wedge dp^1 \wedge dp^2 \wedge dp^3 \qquad (3.154) \\
+ & X_p^1(q,p)\, dq^1 \wedge dq^2 \wedge dq^3 \wedge dp^2 \wedge dp^3 \\
+ & X_p^2(q,p)\, dq^1 \wedge dq^2 \wedge dq^3 \wedge dp^1 \wedge dp^3 \\
+ & X_p^3(q,p)\, dq^1 \wedge dq^2 \wedge dq^3 \wedge dp^1 \wedge dp^2).
\end{aligned}
$$

The pullback $i^*(i_{X_H}\ell)$ maps fibers T_q^*U to fibers $T_q^*\partial U$ and hence for the base the pullback is a mapping $\Omega^k(U) \to \Omega^k(\partial U)$. But the last three components of $i_{X_H}\ell$ in Eq. 3.154 are 3-forms over U and hence vanish when pulled-back onto the 2-dimensional boundary ∂U; the components represent momentum flux in the volume U and hence, to first order, do not contribute to the change in the total phase space density in U. With Remark 2.114 we therefore have

$$
\begin{aligned}
i^*(i_{X_H}\ell) = \mathcal{L}(u,p)(\quad & X_q^1(u,p)\, n_1(u)\, du^1 \wedge du^2 \wedge dp^1 \wedge dp^2 \wedge dp^3 \\
+ & X_q^2(u,p)\, n_2(u)\, du^1 \wedge du^3 \wedge dp^1 \wedge dp^2 \wedge dp^3 \qquad (3.155) \\
+ & X_q^3(u,p)\, n_3(u)\, du^1 \wedge du^2 \wedge dp^1 \wedge dp^2 \wedge dp^3)
\end{aligned}
$$

where $u = (u_1, u_2)$ are local coordinates on ∂U and $\vec{n}(u) = (n_1(u), n_2(u), n_3(u))$ are the components of the local surface normal and which, without loss of generality, is assumed to be of unit length. The fiber integral over $T_q^*\partial U$, whose physical significance arises from Eq. 3.127 and hence the transport theorem, is thus

$$
\begin{aligned}
\int_{T_q^*\partial U} i^*(i_{X_H}\ell) = \quad & J^1(u)\, n_1(u)\, du^1 \wedge du^2 \\
+ & J^2(u)\, n_2(u)\, du^1 \wedge du^3 \qquad (3.156) \\
+ & J^3(u)\, n_3(u)\, du^1 \wedge du^2
\end{aligned}
$$

where the components $J(u) = (J^1(u), J^2(u), J^3(u))$ are given by

$$J^i(u) = \int_{T_q^*\partial U} \mathcal{L}(u,p)\, X_q^i(u,p)\, dp. \qquad (3.157)$$

For pre-radiance density, an analogous derivation can be employed. Using that only the spatial components X_q^i of X_H are relevant for $i^*(i_{X_H} n^3 \tilde{\ell})$, and

that these are not affected by changing the parametrization in the fiber, we obtain[160]

$$
\begin{aligned}
i^*\Big(\mathrm{i}_{X_H}(n^3\,\tilde{\ell})\Big) = n^3\,\tilde{L}(u,\nu,\bar{p})(\quad & X_q^1(u,p)\,n_1(u)\,du^1\wedge du^2\wedge d\nu\wedge d\bar{p} \\
& +X_q^2(u,p)\,n_2(u)\,du^1\wedge du^2\wedge d\nu\wedge d\bar{p} \quad (3.158) \\
& +X_q^3(u,p)\,n_3(u)\,du^1\wedge du^2\wedge d\nu\wedge d\bar{p})
\end{aligned}
$$

where half of the components of the interior product again vanish since they represent an 3-form on the 2-manifold ∂U. Using again the expressions for the configuration space components X_q^i of the Hamiltonian vector field from Eq. 3.40 we obtain

$$
\begin{aligned}
i^*\Big(\mathrm{i}_{X_H}(n^3\,\tilde{\ell})\Big) = c\,n^2\,\tilde{L}(u,\nu,\bar{p})(\quad & \bar{p}_1\,n_1(u)\,du^1\wedge du^2\wedge d\nu\wedge d\bar{p} \\
& +\bar{p}_2\,n_2(u)\,du^1\wedge du^2\wedge d\nu\wedge d\bar{p} \quad (3.159) \\
& +\bar{p}_3\,n_3(u)\,du^1\wedge du^2\wedge d\nu\wedge d\bar{p})
\end{aligned}
$$

where we denoted the components of the unit momentum by $\bar{p} = p/\|p\| = (\bar{p}_1,\bar{p}_2,\bar{p}_3)$. The fiber integral is therefore

$$
\begin{aligned}
\int_{T_q^*\partial U} i^*\Big(\mathrm{i}_{X_H}n^3\tilde{\ell}\Big) = c\,n^2(\quad & \tilde{J}^1(u)\,n_1(u)\,du^1\wedge du^2 \\
& +\tilde{J}^2(u)\,n_2(u)\,du^1\wedge du^3 \quad (3.160) \\
& +\tilde{J}^3(u)\,n_3(u)\,du^1\wedge du^2)
\end{aligned}
$$

with the components $\tilde{J}(u) = (\tilde{J}^1(u), \tilde{J}^2(u), \tilde{J}^3(u))$ with respect to pre-radiance given by

$$
\tilde{J}^i(u) = \int_{S^2}\int_{\mathbb{R}^+} \bar{p}^i\,\tilde{L}(u,\nu,\bar{p})\,d\nu\,d\bar{p}. \tag{3.161}
$$

When the normal is kept as part of the integrand one has for the fiber integral

$$
\int_{S^2}\int_{\mathbb{R}^+} c\,n^2\,i^*\Big(\mathrm{i}_{X_H}\tilde{\ell}\Big) = c\,n^2(q)\left(\int_{S^2}\int_{\mathbb{R}^+} \tilde{L}(u,\nu,\bar{p})(\bar{p}\cdot\vec{n})\,d\nu\,d\bar{p}\right)dA. \tag{3.162}
$$

As discussed in detail before, classical radiance is obtained when one considers the integrand in the above equation before it is pulled back onto a surface.

Remark 3.28. In the foregoing derivations, we required both the paracompactness of T^*U and the pullback by the inclusion map $i:\partial W \to W$ for Stokes' theorem to be applicable and to simplify the resulting surface integral. Hence, a surprising technical sophistication was required to derive the measurement

---

[160]An obvious argument for the invariance of the components X_q^i is covariance which is required by physics.

equation. It should also be noted that in the derivation for pre-radiance, the Hamiltonian vector field led "serendipitously" to both the quadratic dependence that is known from the literature, by cancelling one factor of the cubic dependence of pre-radiance, and to the cosine term $(\vec{n} \cdot \bar{p})$ with the unit vector $\bar{p}$, which is already present in the original form of Hamilton's equations in Eq. 3.40.

3.2.7 Scattering

Scattering arises microscopically from the interaction of the electromagnetic field with non-conducting regions in the transport medium. A solution of Maxwell's equations under appropriate boundary conditions is then required to characterize the propagation of the electromagnetic field and energy density.[161] However, when the transport of the electromagnetic energy density is modelled on phase space T^*Q, the macroscopic effects of the interaction can be described by effective mathematical models,[162] and these will be the subject of the present section.[163]

3.2.7.1 Scattering at Surfaces

Scattering at surfaces can be understood as the re-emission of an energy density flow impinging onto a surface. It can hence be modelled analogously to the measurement of light energy density discussed in the previous section.

Surface scattering is macroscopically described by the **scattering equation**

$$\hat{\ell} = \int_{T^-\mathcal{M}} \star \left(i_{X_H} \ell \right) \wedge \rho \tag{3.163}$$

where $\hat{\ell}(\hat{q}, \hat{p})$ is the phase space energy density after scattering and $i_{X_H}\ell$ represents the impinging flow on the surface analogous to Eq. 3.134. The scattering kernel is an asymmetric differential double form, cf Remark 2.98, given by

$$\rho = P(u, p, \hat{q}, \hat{p})\, du \wedge dp \otimes d\hat{q} \wedge d\hat{p} \in \Omega^5(T^-\mathcal{M}) \times \mathrm{Den}(T^*\mathcal{M}) \tag{3.164}$$

---

[161]We would like to mention that popular scattering models that employ electromagnetic theory but model the surface as a collection of microfacets are *not* physical in a rigorous sense since these are not approximations or asymptotic limits of a description using more accurate and realistic models for the boundary surfaces. Classical examples of microfacet models are the Torrance-Sparrow model ("Theory for Off-Specular Reflection From Roughened Surfaces") and its adaptation by Cook and Torrance ("A Reflectance Model for Computer Graphics"), and the model by He at al. ("A Comprehensive Physical Model for Light Reflection"), see also the classic text (Beckmann and Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*).

[162]This advantage of a description on phase space was explicitly pointed out in (Ryzhik, Papanicolaou, and Keller, "Transport equations for elastic and other waves in random media").

[163]A detailed discussion on the transition from microscopic to macroscopic models for scattering can be found in (Mishchenko, Travis, and Lacis, *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering*; Mishchenko, "Radiative Transfer: A New Look of the Old Theory") and references therein. Ryzhik, Papanicolaou, and Keller ("Transport equations for elastic and other waves in random media") provide a derivation of scattering that arises from inhomogenities in the refractive index at length scales which are at the order of the wavelength.

where $du = du^1 \wedge du^2$ is the area form for the local coordinates $u = (u_1, u_2)$ on the 2-manifold $\mathcal{M} \subset \mathbb{R}^3$, $T^-\mathcal{M}$ is again the negative half-space of the restriction $T^*\mathcal{M} = T^*Q \mid_{\mathcal{M}}$, see Chapter 3.2.6.2, and the Liouville form $d\hat{q} \wedge d\hat{p}$ has support over the phase space point $(\hat{q}, \hat{p})$ after scattering. The Hodge dual in Eq. 3.163 is with respect to the induced Riemannian structure on the surface $\mathcal{M}$, and it thus takes the form

$$\star \left(i_{X_H} \ell \right) = \star \left(\frac{c}{n(q)} F(u, p) \, (\bar{p} \cdot \vec{n}) \, dp \, dA \right) \tag{3.165a}$$

$$= \frac{c}{n(q)} F(u, p) \, (\bar{p} \cdot \vec{n}), \tag{3.165b}$$

cf. Eq. 3.134c. Hence, analogous to the measurement equation in Eq. 3.134c, the scattering equation can be written as

$$\hat{\ell} = \left(\int_{T^-\mathcal{M}} \frac{c}{n(q)} F(u, p) \, P(u, p, \hat{q}, \hat{p}) \, (\bar{p} \cdot \vec{n}) \, dp \, dA \right) d\hat{q} \wedge d\hat{p} \tag{3.166}$$

where $c/n(q)$ again represents the local flux rate at $q \in \mathcal{M}$. The above scattering function is non-local on the surface and it also admits fluorescence, that is the re-emission of light at a wavelength different than those of the incoming radiation. It hence corresponds to a generalized bidirectional (sub-)surface scattering distribution function in classical nomenclature,[164] and this is mathematically the natural mode of scattering associated with our model of light transport theory. Parametrizing the scattering kernel in spherical coordinates and with respect to frequency we obtain

$$\tilde{\rho} = n^3(u) \, n^3(\hat{q}) \, \tilde{P}(u, \bar{p}, \nu, \hat{q}, \hat{\bar{p}}, \hat{\nu}) \, du \, d\bar{p} \, d\nu \otimes d\hat{q} \, d\hat{\bar{p}} \, d\hat{\nu} \tag{3.167}$$

where constant factors have already been subsumed into the coordinate function $\tilde{P}(u, \bar{p}, \nu, \hat{q}, \hat{\bar{p}}, \hat{\nu})$, see the change of variables leading from the phase space energy density to pre-radiance in Chapter 3.2.5. The Hodge dual of the interior product in Eq. 3.165b becomes in spherical coordinates

$$\star_s \left(i_{X_H} (\eta^3 \tilde{\ell}) \right) = \star_s \left(c \, n^2(q) \, \tilde{L}(u, \bar{p}, \nu) (\bar{p} \cdot \vec{n}) \, d\bar{p} \, d\nu \, dA \right) \tag{3.168a}$$

$$= \frac{c}{n(q)} \tilde{L}(u, \bar{p}, \nu) (\bar{p} \cdot \vec{n}) \tag{3.168b}$$

where the factor $1/n^3(q)$ arises through the Hodge star $\star_s$ for the induced metric in spherical coordinates, cf. Remark 2.90. With the re-parametrized scattering kernel, Eq. 3.166 thus becomes

$$\frac{\hat{\ell}}{n^3(\hat{q})} = \left(\int_{\mathbb{H}^-\mathcal{M}} \int_{\mathbb{R}^+} c \, n^2(u) \, \tilde{L}(u, \bar{p}, \nu) \, \tilde{P}(u, \bar{p}, \nu, \hat{q}, \hat{\bar{p}}, \hat{\nu}) \, (\bar{p} \cdot \vec{n}) \, dp \, dA \right) d\hat{q} \, d\bar{p} \, d\nu$$

[164]Nicodemus and Kostkowski, *Self Study Manual on Optical Radiation Measurements*.

and using the definition of pre-radiance from Chapter 3.2.5 this is equivalent to

$$\hat{\ell} = \left(\int_{\mathbb{H}^-} \int_{\mathcal{M}} \int_{\mathbb{R}^+} c\, n^2(u)\, \tilde{L}(u, \bar{p}, \nu)\, \tilde{P}(u, \bar{p}, \nu, \hat{q}, \hat{\bar{p}}, \hat{\nu})\, (\bar{p}{\cdot}\vec{n})\, dp\, dA \right) d\hat{q}\, d\bar{p}\, d\nu. \quad (3.169)$$

The scattering model considered so far is in practice usually intractable. Hence, some or most of the dimensions are typically assumed to be trivial. For example, assuming the scattering to be local at $u = \hat{q}$ and preserving the frequency so that $\nu = \hat{\nu}$ we obtain for Eq. 3.169 that

$$\hat{\ell} = c\, n^2(u) \left(\int_{\mathbb{H}^-} \tilde{L}(u, \bar{p}, \nu)\, \tilde{P}(u; \bar{p}, \hat{\bar{p}})\, (\bar{p} \cdot \vec{n})\, dp \right) d\hat{q}\, d\bar{p}\, d\nu \quad (3.170)$$

and $\tilde{P}(u; \bar{p}, \hat{\bar{p}})$ is then the equivalent of the classic bidirectional reflectance distribution function. Eq. 3.170 is known as the shading equation and in the classical literature often considered as the central equation in light transport theory in environments where the refractive index is homogeneous.[165]

As discussed in the foregoing, in the literature attempts have been made to derive scattering functions from Maxwell's equations. In contrast, we consider them as integral boundary condition[166] that are provided as input to a transport problem. In the light of our previous formulation of light transport, it is thereby reasonable to require the scattering kernel to be symmetric and positive in the momentum,

$$P_{u,\bar{q}}(p, \bar{p}) = P_{u,\bar{q}}(\bar{p}, p) \qquad \text{(Helmholtz reciprocity)} \qquad (3.171a)$$

$$P_{u,\bar{q}}(p, \bar{p}) \geq 0 \qquad \text{(positivity)} \qquad (3.171b)$$

The second requirement follows immediately from the physical significance of the phase space light energy density. Helmholtz reciprocity is natural for a Hamiltonian system which is by construction time-reversible although its justification in general contexts can be difficult.[167] We also require the reciprocity because it is critical for most computational techniques that are currently employed.[168]

---

[165]See for example (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*, p. 296), the equation is sometimes also denoted as reflectance equation.

[166]A similar approach was taken for example by Arvo ("Transfer Equations in Global Illumination").

[167]For modern discussions of Helmholtz reciprocity see for example (Clarke and Parry, "Helmholtz Reciprocity: Its Validity and Application to Reflectometry"; Snyder, "Reciprocity of the bidirectional reflectance distribution function (BRDF) in measurements and models of structured surfaces"). For some experimental evaluation of the reciprocity and its failure (Okayama and Ogura, "Experimental verification of nonreciprocal response in light scattering from rough surfaces"; Venable, "Comments on reciprocity failure"; Kim, "Verification of the reciprocity theorem"; Kriebel, "On the Limited Validity of Reciprocity in Measured BRDFs").

[168]It is Helmholtz reciprocity that make the light transport operator self-adjoint and hence allows to trace rays from the camera instead of the light as required by physics.

3.2.7.2 Volume Scattering

The time evolution equation for light transport when volume scattering is included is the inhomogeneous transport equation

$$\frac{\partial \ell}{\partial t} + \{\ell, H\} = \left. \frac{\partial \ell}{\partial t} \right|_{\text{scatt.}} \tag{3.172}$$

where the right hand side models scattering. With the usual scattering assumptions,[169] and using the form of the light transport equation in Eq. 3.50b, one obtains

$$\frac{\partial \ell}{\partial t} - \pounds_{X_H} \ell = \sigma \pounds_{X_H} \ell + \int_{T_q^* Q} \star_p \ell \wedge \varrho \tag{3.173}$$

where $\star_p$ represents the Hodge dual of the momentum variable with respect to the canonical Euclidean structure on the fibers $T_q^* Q \cong \mathbb{R}^3$. The volume scattering kernel

$$\varrho = P(q, p, \bar{p}) \, dp \otimes d\bar{p} \in \Omega^3(T^* Q) \otimes \Omega^3(T^* Q) \tag{3.174}$$

is again a differential double form[170] and it satisfies Helmholtz reciprocity and positivity as defined in Eq. 3.171. The scattering coefficient $\sigma = \sigma_a + \sigma_o$ combines the absorption coefficient σ_a, which taken alone represents the Beer-Lambert law, and the out-scattering coefficient σ_s which is defined by the fiber integral

$$\sigma_s = \int_{T_q^* Q} \star_p \varrho \tag{3.175}$$

and the integral yields a section of $\mathcal{F}(Q)$ as required by Eq. 3.173.[171]

Remark 3.29. As discussed in Remark 3.2 and as is also apparent from the discussion in Chapter 3.2.7.1, scattering at surfaces does not break the Hamiltonian structure of light transport but it restricts the flow domain. In contrast, volume scattering as described by Eq. 3.173 destroys the Hamiltonian character. However, when only absorption along the flow of the system is considered, which is classically described by the Beer-Lambert law, then the volume transport equation takes the form

$$0 = \frac{\partial \ell}{\partial t} - \pounds_{X_H} \ell - \sigma_a \pounds_{X_H} \ell \tag{3.176a}$$

---

[169]See for example (Arvo, "Transfer Equations in Global Illumination"; Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*)

[170]The volume scattering kernel is a degenerate 3-form over phase space in that only the basis function coefficient with respect to $dp = dp^1 \wedge dp^2 \wedge dp^3$ is nonzero; it hence has a structure similar to the canonical 1-form. In each fiber, ϱ is a volume double form.

[171]That the Lie derivative $\sigma \pounds_{X_H} \ell$ is equivalent to the classical formulation "σL", which should really be written as "$\sigma L \, ds$", is clear since $\pounds_{X_H} \ell$ and $L \, ds$ both represent the Lagrangian rate of change for a conservative system.

$$= \frac{\partial \ell}{\partial t} - (1 + \sigma_a) \pounds_{X_H} \ell \tag{3.176b}$$

$$= \frac{\partial \ell}{\partial t} - \beta \pounds_{X_H} \ell \tag{3.176c}$$

and since the light energy density is only attenuated along trajectories in phase space the Hamiltonian structure of light transport is retained. One can consider β as a drag coefficient, although it is important to keep in mind that it is a phase space function since $\sigma_a \in \mathcal{F}(T^*Q)$ and the absorption can vary with location, direction, and frequency.[172]

Eq. 3.176 describes a linear dissipative system where energy is dissipated but the flow in phase space is preserved. This is possible since the phase space energy density ℓ is transported or "Lie-dragged" by X_H but the Hamiltonian vector field X_H does not depend on the density ℓ. An analogous model for dissipation that preserves the Hamiltonian structure but decays energy is provided by the consecutive application of two Poisson brackets. Such double brackets arose independently in the context of plasma physics,[173] astrophysics,[174] fluid dynamics,[175] and linear algebra,[176] and were later recognized as a general principle for dissipation in Hamiltonian systems.[177] Recently, they have received considerable attention in the context of kinetic theory and its approximation using kinetic and statistical moments,[178] cf. Remark 3.9, and they are also of practical importance for climate modeling.[179]

Remark 3.30. Modeling of systems with scattering or nonlinear dissipation, which includes the Navier-Stokes equations, is one of the major open problems in geometric mechanics.[180] It has to be noted that much of the elegant theory

---

[172]The difference between a coefficient and a phase space function becomes particularly clear when one considers Eq. 3.176c in Poisson bracket form. For a scalar one has by bilinearity that $\beta\{\ell, H\} = \{\ell, \beta H\}$ whereas for a phase space function the Leibniz rule has to be employed and one has $\beta\{\ell, H\} = \{\ell, \beta H\} - \{\ell, \beta\}H$.

[173]Turski and Kaufman, "Canonical-dissipative formulation of relativistic plasma kinetic theory with self-consistent Maxwell field".

[174]Kandrup, "The secular instability of axisymmetric collisionless star clusters".

[175]Vallis, Carnevale, and Young, "Extremal energy properties and construction of stable solutions of the Euler equations".

[176]Brockett, "Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems"; Brockett, "Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems".

[177]Bloch et al., "Dissipation Induced Instabilities"; Otto, "The Geometry of Dissipative Evolution Equations: The Porous Medium Equation".

[178]Holm, Putkaradze, and Tronci, "Geometric dissipation in kinetic equations"; Morrison, "Thoughts on brackets and dissipation: Old and new"; Holm, Putkaradze, and Tronci, "Kinetic Models of Heterogeneous Dissipation"; Holm, Putkaradze, and Tronci, "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions".

[179]In two-dimensional fluids, the dissipation modeled using double brackets leads to an inverse energy cascade and this has important consequence for the dynamics of the atmosphere, cf. (Holm, Putkaradze, and Tronci, "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions") and references therein.

[180]For some recent work on the problem see for example (Cruzeiro, "Hydrodynamics, probability and the geometry of the diffeomorphisms group"; Gliklikh, "Deterministic viscous

can be preserved when only perfectly specular scattering is allowed, and this is studied for example in billiard theory.[181]

3.2.8   Discussion

In the following, we will put the foregoing technical considerations into perspective, relate them to previous work in the literature, and discuss possible directions for future work.

From Maxwell's Equations to Light Transport   In this chapter, we illuminated the physics and mathematics of light transport by deriving the governing equations from Maxwell's equations and studying its structure using geometric mechanics, with the use of modern tensor calculus being vital to rigorously and transparently work on the five and six dimensional phase space over which the theory is naturally defined, and which is beyond the confines of classical vector calculus that serves so well on configuration space.

We employed microlocal and semi-classical analysis, central tools in the modern theory of partial differential equations, to lift Maxwell's equations from configuration to phase space. This provided an equivalent description of electromagnetic theory on the cotangent bundle where the electromagnetic field is represented by a matrix-valued Wigner distribution. We showed that the short wavelength limit on phase space gives rise to Hamiltonian dynamics with the electromagnetic energy then being described by a phase space light energy density that represents unpolarized radiation. The Hamiltonian function of light transport that is obtained from the derivation is homogeneous of degree one, and we showed that the associate $\mathbb{R}^+$-symmetry, an imprint from Maxwell's equations at the short wavelength limit, enables to reduce the dynamics from the six dimensional slit cotangent bundle $T^*Q \setminus \{0\}$ to the five dimensional cosphere bundle $S^*Q = (T^*Q \setminus \{0\})/\mathbb{R}^+$. Using a generalized Legendre transform, we established the equivalence of our Hamiltonian formulation of light transport and Fermat's principle, and this also showed that time evolution is described by a geodesic flow along a spatially inhomogeneous metric that depends quadratically on the refractive index.

The central quantity in our formulation of light transport is the phase space light energy density $\ell_t \in \mathrm{Den}(T^*Q)$ that provides a representation of the electromagnetic energy density $\mathcal{E}(q,t)$ on the cotangent bundle, and which can

---

hydrodynamics via stochastic processes on groups of diffeomorphisms"; Arnaudon, Cruzeiro, and Galamba, "Lagrangian Navier-Stokes flows: a stochastic model"; Cipriano and Cruzeiro, "Navier-Stokes Equation and Diffusions on the Group of Homeomorphisms of the Torus"; Watanabe, "Differential geometry on diffeomorphism groups and Lagrangian stability of viscous flows").

[181]See (Smilie, "The Dynamics of Billiard Flows in Rational Polygons") for a recent survey. The measure-valued flows considered by Brenier ("The Least Action Principle and the Related Concept of Generalized Flows for Incompressible Perfect Fluids") provide a possible alternative, but much of the structure is lost with this approach.

be recovered from ℓ_t by the fiber integral

$$\mathcal{E}(q,t) = \int_{T_q^*Q} \ell_t \, dp.$$

Time evolution of the light energy density ℓ_t is governed by the canonical Poisson bracket $\dot{\ell}_t = \{\ell_t, H\}$ and an explicit transport equation, paralleling those in the classical literature, is obtained with the coordinate expression of the bracket and Hamilton's equations. An equivalent, finite time description of the propagation of ℓ_t is provided by the pullback $\eta_t^* \ell_t = \ell_0$ along the Hamiltonian flow $\eta_t : \mathbb{R} \times T^*Q \to T^*Q$, which provides an intrinsic formulation for the ray casting function found in the classical literature. With a change of variables to spherical coordinates in the fibers of T^*Q, as dictated by the tensorial nature of the light energy density, we obtained pre-radiance $\tilde{\ell}$, which enabled to recover classical radiometry when measurements are considered. Defining pre-radiance $\tilde{\ell}$ to be consistent with traditional radiance provokes however considerable difficulties when transport is considered, since only the light energy density ℓ but not pre-radiance $\tilde{\ell}$ is conserved along trajectories in phase space. In homogeneous media without spatial variations in the refractive index, these differences disappear, explaining why historically pre-radiance and radiance are considered as central quantities in light transport theory.

Studying the transport theorem of continuum mechanics for light transport theory enabled us to understand how macroscopic observables can be obtained from the microscopic phase space light energy density. Moreover, separating surface independent terms from measurable quantities connected our formulation of light transport to the phenomenological theory of radiometry, with for example classical radiance being obtained as a 2-form that represents measurements of pre-radiance through an arbitrary surface. The pullback of the 2-form then yields the cosine or foreshortening factor that is omnipresent in classical radiometry, but which to our knowledge has previously only been justified using heuristic, pre-Descartian arguments. A corollary of these results is that classical radiometry is inherently associated with measurements, a fact which has recently often been overlooked in the literature, while what is transported is the phase space energy density and pre-radiance—a vital distinction which can only be fully understood with our formulation of light transport theory using modern tensor calculus.

Our derivation of light transport theory using microlocal and semi-classical analysis is geometric in nature, going from the geometric Maxwell's equation to a geometric Hamiltonian system, and it exposes a rich fabric that underlies and rationalizes the passage. Critical for the derivation and for obtaining the geometric structure was to consider inhomogeneous media with spatially varying refractive index, since in the homogeneous case, which is more commonly encountered in applications, the structure of the theory is obfuscated. Our ansatz using semi-classical analysis also reveals remarkable parallels to other important questions in mathematical physics, such as (inverse) quantization, and these correspondences provide important additional insights. The exclusion of volume scattering was a necessary assumption for our theoretical investigations,

although it is rightly considered as a major limitation of the current results. Scattering at surfaces, in contrast, is easily incorporated into our formulation by restricting the Hamiltonian flow to be local in between surfaces. The rich structure that arises when the Hamiltonian flow is defined globally will be studied in the next section.

For us, a surprising outcome of our investigations was the rigorous physical justification of many of the intuitive connections between Maxwell's equations and classical radiometry that have been considered in the past, such as the identification of vector irradiance with the magnitude of the Poynting vector or the equivalence of the intensity law of geometric optics with the light transport equation. However, our derivation shows that these connections do not arise serendipitously, but that they emerge because light transport is obtained from an alternative *but equivalent* representation of electromagnetic theory on phase space—what is surprising, nonetheless, is that the equivalence also holds at the short wavelength limit. To our knowledge, not known before was the equivalence of the light transport equation on phase space and the transport equation in Poynting's theorem on configuration space. Although in its current form the result is restricted to domains where the boundary integral of the light energy density vanishes,[182] it will be interesting to explore which additional terms appear for general domains.

We believe that the geometric structure of light transport which we developed in this section provides a principally complete picture of the fabric underlying the theory. Some details still require refinement, see the ensuing discussion, but these will in our opinion support the structure without affecting its shape.

Position within the Literature   As we discussed in detail in the foregoing, a considerable amount of the material presented in the previous sections appeared before in the literature. However, caused by the fragmentation of radiometry in the 20th century, the results are scattered throughout various research communities, and many connections and aspects have not been appreciated before.

In plasma physics, the Hamiltonian structure of radiative transfer and its derivation from Maxwell's equations using the Wigner transform are known since the 1960s. Nonetheless, in many areas such as computer graphics the theory is still justified based on transport theory and using photon counts, or with phenomenological considerations. In theoretical optics, the connection between classical electromagnetic theory and radiometry were investigated since the 1970s. However, the missing understanding of radiometric concepts and the lacking usage of modern tensor calculus led to limited insight and a restricted applicability of the obtained results, and subsequently this also affected work in other fields, such as computer vision and computational photography, where

---

[182]The equivalence appeared in (Ryzhik, Papanicolaou, and Keller, "Transport equations for elastic and other waves in random media") although these authors did not discuss possible restrictions or implications, and it does not appear to be known elsewhere in the literature.

the approach has been employed.[183] In geometric mechanics, the mathematical structure of ray optics and Fermat's principle are well known. The Hamiltonian formulation of classical optics, however, was to our knowledge only investigated in the Russian literature, and only systems with additional symmetries such as fiber optics were considered in the West. Additionally, neither community employed the results in the context of light transport theory to describe the propagation of the phase space light energy density, although various aspects have been studied unwittingly in the context of (inverse) quantization and other continuum theories. The derivation of light transport theory from Maxwell's equations that we outlined in the foregoing was obtained in applied mathematics as a variation of the semi-classical limit of the Schrödinger equation. However, the geometric structure that is apparent in the ansatz was not investigated in this community and parallels to geometric quantization were not considered or mentioned. Lacking there was also a tensorial characterization of concepts such as the Wigner transform.

For our presentation, we synthesized material from various communities and formulated it using the modern language of geometric mechanics and tensor calculus, providing a vigorous mathematical foundation for light transport that enables to rigorously study the theory. Our analysis of the structure of light transport led to some corrections of results that appeared before in the literature, and to our knowledge we are the first to rigorously relate classical radiometry to Maxwell's equations. The literature on light transport in computer graphics is unfortunately devoid of reference to the presented material, and there the theory is still largely at the point which was characterized as "arrested development" already in the 1930s.[184] As discussed in Chapter 1.1.2, in the 1990s some attempts were made in the computer graphics community to put the formulation of light transport employed there on a more rigorous and insightful mathematical basis. However, the measure theoretic concepts that were employed lack the smoothness which gives our formulation most of its power, while the functional analytic aspects were never considered in sufficient detail to yield practical benefits.

Future Work   The geometric structure of light transport and its physical foundations developed by us provide many avenues for future work. For us, the most interesting ones are a complete geometrization of the derivation of light transport from Maxwell's equations, and an application of the geometric structure and the insights provided by it to develop more effective computational techniques.

Our derivation of light transport theory from Maxwell's equation followed the recent literature in applied mathematics. It is hence currently not geometric, even though such a structure is already apparent underneath the computations. We believe that a geometrization based on the space-time formulation of

---

[183]A singular exception where the Wigner transform has been employed in the context of light transport simulation, inspired by the work in theoretical optics, is (Oh et al., "Rendering Wave Effects with Augmented Light Field").

[184]Preface to (Gershun, "The Light Field").

Maxwell's equations will provide further insight into the passage from configuration to phase space and lead to an intrinsic formulation using tensor notation. Inspiration for this endeavour is provided by the modern theory of first order partial differential equations and the close analogy between the derivation employed by us and inverse quantization, which has been studied extensively from a geometric point of view.[185] Interesting questions which should thereby be considered are for example if an analogue to the Moyal bracket for optical coherence theory exists and if the phase space coherence density is related to the symmetric tensors obtained in recent work on kinetic and statistical moments.[186] For the Wigner transform, it would also be interesting to study it from a representation theoretic point of view, and to investigate possible connections to the plasma-to-fluid map in geometric mechanics.[187] A geometrized derivation of light transport theory should also provide a connection to the recent work by Mishchenko for light transport in media with scattering particles. This would provide a more complete picture of the physical foundations of the theory, and extend its range of validity.[188]

Microlocal and semi-classical analysis, which provide the language for the derivation of light transport theory from Maxwell's equations, are closely related to Fourier integral operators that are endowed with a well known infinite dimensional Lie group structure.[189] It is tantalizing to assume that the group of symplectic transformations that underlies light transport, and which will be studied in detail in the next section, arises from this group structure of Fourier integral operators. Similar to the transition from Maxwell's equations to light transport, the BBGKY hierarchy in classical kinetic theory enables to

---

[185]See the discussion in (Arnold, *Lectures on Partial Differential Equations*), (Littlejohn and Flynn, "Phase integral theory, coupled wave equations, and mode conversion."; Emmrich and Weinstein, "Geometry of the transport equation in multicomponent WKB approximations") and in particular (Guillemin and Sternberg, *Semi-Classical Analysis*) for existing work.

[186]Gibbons, Holm, and Tronci, "Geometry of Vlasov kinetic moments: A bosonic Fock space for the symmetric Schouten bracket".

[187]Marsden and Ratiu, *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*, p. 366.

[188]A partial comparison between continuum and discrete scatterer approaches is available in the book by Mishchenko (*Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering*), see in particular the introduction to Chapter 8 for an overview of early attempts for a derivation of radiative transfer from Maxwell's equations, and the opinion articles (Mishchenko, "Radiative Transfer: A New Look of the Old Theory"; Mishchenko, "Gustav Mie and the fundamental concept of electromagnetic scattering by particles: A perspective"; Mishchenko et al., "Electromagnetic scattering by a morphologically complex object: Fundamental concepts and common misconceptions") and references therein.

[189]See (Ratiu and Schmid, "The differentiable structure of three remarkable diffeomorphism groups"; Adams, Ratiu, and Schmid, "A Lie group structure for pseudodifferential operators"; Adams, Ratiu, and Schmid, *A Lie group structure for Fourier integral operators*), and (Khesin and Wendt, *The Geometry of Infinite-Dimensional Groups*) for a recent discussion and it should be noted that Fourier integral operators include pseudo-differential operators (Wunsch, *Microlocal Analysis and Evolution Equations*, p.60). Fourier integral operators have been studied extensively from the group point of view since they appear naturally as Lie algebra in the Korteweg-de Vries equation and provide an explanation of its integrability (Adler, "On a trace functional for formal pseudo-differential operators and the symplectic structure of the Korteweg-deVries type equations").

obtain a macroscopic Hamiltonian system from a microscopic description of the dynamics. It would be interesting to explore if an intrinsic connection and a common geometric structure for the approaches exists.[190]

Some additional theoretical questions which appear interesting to us are concerned with a continuation of the analysis of the geometric structure of light transport. In the Legendre transform, we distinguished between a canonical and a kinetic momentum, with the latter one being identified with the wave vector. Conceptually, this appears to be the correct description, but a more formal comparison and analysis to the analogous concepts in electromagnetic theory is required. Also in the context of the Legendre transform, we showed that the natural metric for light transport depends on the refractive index and that time evolution can be considered as a geodesic flow for this Riemannian structure. In subsequent considerations, however, we did not employ the metric. We believe that doing so would have led to simplifications in particular in the context of pre-radiance.

Throughout the previous sections, we related light transport theory to classical radiometry and optics, and we recovered many known laws from these fields. Our conclusion and results on this matter should however be considered with some caution since an interpretation of the classical literature is required to relate it to modern, rigorous concepts. For example, our interpretation of the intensity law of geometric optics does not fully agree with the intuitive picture one has in mind, where intensity is associated with a surface transported through configuration space. Further work, and also a more detailed study of how the laws and concepts arose historically, seems necessary to solidify our interpretations.

Next to the many theoretical questions discussed in the foregoing, our modern, geometric formulation of light transport also opens up many possibilities for applications. The modern mathematical structure developed by us enables to couple light transport to other dynamical systems such as fluids. This is vital for example in climate modeling, where the atmosphere can be described by a fluid but much of the dynamics are driven by radiation from the sun. A coupling to other mechanical systems is also necessary for an evolution operator: an algorithm that enables to incrementally update a sequence of images from frame to frame instead of computing every image independently, and which saves computations by doing so. With the equivalence between the light transport equation and Poynting's theorem which we established, one is tempted to replace the flow on phase space by Poynting's transport equation on the lower dimensional configuration space.[191] This would break the "curse of dimensionality" for light transport, and make for example finite element methods again a viable option.[192] As mentioned before, the equivalence only holds when

---

[190]The geometry of the BBGKY hierarchy was studied in (Marsden, Morrison, and Weinstein, "The Hamiltonian structure of the BBGKY hierarchy equations").

[191]This is in line with some recent work in the literature (Benamou, "Direct computation of multivalued phase space solutions for Hamilton-Jacobi equations").

[192]It has to be noted that multi-grid methods enable to efficiently employ finite element methods on a five or six dimensional domain, see for example (Bungartz and Griebel, "Sparse

the boundary integral of the phase space light energy density vanishes, although work in fluid dynamics suggests that computationally one might be able to relax this restriction.[193] Another caveat of a simulation on configuration space would be that the angular dependence is lost, which currently is vital for the representation of scattering effects. However, we believe that it will be possible to employ the Wigner transform after transport to reconstruct a phase space representation.

We discussed previously that the Beer-Lambert law and double Poisson brackets provide mechanisms for dissipation that preserve the Hamiltonian structure of a system, and hence the advantages of its rich, geometric theory. Nonlinear double Poisson brackets provide considerable flexibility in the effects that can be modelled, and we believe that these enable to obtain light scattering models which are realistic enough to be relevant for applications.

An important advantage of a geometric formulation of light transport is that it provides a description independent of many aspects of the underlying space, such as its dimensionality or the metric defined on it. For example, radiation hydrodynamics,[194] which is concerned with light transport at astrophysical length scales, requires a formulation on curved space-time, while in computer graphics often light transport in two dimensional flatland is considered.[195] Exploring the form of our theory in these settings seems both practically relevant and theoretically interesting.

The above mentioned applications require significant changes to the current practice, and have therefore to be considered as long term objectives. However, our results are also of great relevance within the existing state-of-the-art. Recently, light transport in media with varying refractive index has received considerable attention in the literature.[196] The transport equation derived by us provides a firm physical basis for this work, and it appears that some corrections to current practice are necessary since there radiance and the intensity of geometric optics have been employed without a change of variables as required for tensors. Moment approximations for radiance have been employed

---

grids").

[193]Mullen et al., "Energy-Preserving Integrators for Fluid Animation".

[194]See for example (Mihalas and Weibel Mihalas, *Foundations of Radiation Hydrodynamics*; Castor, *Radiation Hydrodynamics*).

[195]Classic references are (Edelsbrunner, Overmars, and Wood, "Graphics in Flatland: A Case Study"; Pocchiola, "Graphics in Flatland Revisited"; Heckbert, "Radiosity in Flatland"; Orti et al., "Radiosity for dynamic scenes in flatland with the visibility complex"; Durand, Drettakis, and Puech, "The 3D Visibility Complex: A New Approach to the Problems of Accurate Visibility") and the ideas were recently revived for example in (Durand et al., "A Frequency Analysis of Light Transport"; Ramamoorthi, Mahajan, and Belhumeur, "A First-Order Analysis of Lighting, Shading, and Shadows"). In the case of flatland, we believe that the difference between two and three dimensions should be far less grave than for fluid dynamics, see for example (Holm, Putkaradze, and Tronci, "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions", 1.a.ii) or (Khesin, "Topological Fluid Dynamics"), but a rigorous study is needed.

[196]Ihrke et al., "Eikonal Rendering: Efficient Light Transport in Refractive Objects"; Atcheson et al., "Time-Resolved 3D Capture of Non-Stationary Gas Flows"; Ihrke et al., "Tomographic Reconstruction and Efficient Rendering of Refractive Gas Flows".

in various forms in the literature.[197] With our formulation of light transport theory these could be extended beyond the first order models that have usually been considered, and one could employ the insight obtained in recent work on moment representations in geometric mechanics.

Appendix A. The Wigner Transform

The Wigner transform[198] is an important aspect of our derivation of light transport theory from Maxwell's equations, and it enables to "lift" the electromagnetic field from configuration space to an isomorphic representation on phase space. The correspondence that the transform establishes between the spaces is intrinsically explained by representation theory, where the Wigner transform and the Weyl map are inverse maps between different representations of a Lie algebra, see again Fig. 3.3.[199] For us, the representation theoretic character of the Wigner transform is important because it shows that the lift from configuration to phase space—and hence also the transition from Maxwell's equations to light transport theory—is geometric in nature. However, we will forego the associated technicalities in the following, which would lead us too far afield, and instead present a pragmatic approach to the Wigner transform that provides its elementary definitions and its relationship to quadratic observables, which is the most important aspect for our purposes.[200]

Let the Fourier transform for functions $F \in \mathcal{F}(Q)$ be given by

$$\hat{F}(p) = \int e^{-ip \cdot q} F(q) \, dq \tag{3.177a}$$

and let the inverse transform be

$$F(q) = \frac{1}{(2\pi)^3} \int e^{ip \cdot q} \hat{F}(p) \, dp. \tag{3.177b}$$

For a density on configuration space

$$\varphi = F(q) \, dq \in \mathrm{Den}(Q) \tag{3.178}$$

---

[197]Kajiya and Herzen, "Ray Tracing Volume Densities"; Stam, "Multiple scattering as a diffusion process"; Arvo, "The Irradiance Jacobian for Partially Occluded Polyhedral Sources"; Arvo, "Applications of Irradiance Tensors to the Simulation of Non-Lambertian Phenomena".

[198]The Wigner transform was introduced by Wigner in ("On the Quantum Correction For Thermodynamic Equilibrium").

[199]The Weyl map is sometimes also known as Weyl correspondence or Weyl quantization. The equivalence established by the Wigner transform and the Weyl map is somewhat at odds with Weyl's original intention ("Quantenmechanik und Gruppentheorie") in his work on the subject where he sought to bridge the gap between quantum mechanics and classical mechanics and hence was interested in the asymptotic limit directly. In the context of quantization the underlying structure is the Weyl algebra.

[200]Our treatment is based on (Gosson, *Symplectic Methods in Harmonic Analysis and in Mathematical Physics*, Part II, Chapter 9 and 19) as well as (Papanicolaou and Ryzhik, "Waves and Transport"; Gérard et al., "Homogenization limits and Wigner transforms").

with $\star\varphi = F(q) \in L_2(Q)$,[201] the **Wigner transform** $W : \mathrm{Den}(Q) \to \mathrm{Den}(T^*Q)$ is defined by

$$W(\varphi) = w_\varphi = W_\varphi(q,p)\, dq \wedge dp \in \mathrm{Den}(T^*Q) \qquad (3.179)$$

where $\varpi = dq \wedge dp$ is the usual Liouville form on the cotangent bundle T^*Q, and the coordinate function W_φ of the Wigner quasi probability distribution w_φ is

$$(W\varphi)(q,p) = W_\varphi(q,p) = \frac{1}{(2\pi)^3} \int_Q e^{ip\cdot r}\, F\left(q - \frac{1}{2}r\right) F\left(q + \frac{1}{2}r\right) dr. \qquad (3.180)$$

The part of Eq. 3.180 that depends on F is sometimes referred to as the mutual intensity function

$$J(q,r) = F\left(q - \frac{1}{2}r\right) F\left(q + \frac{1}{2}r\right) \qquad (3.181)$$

and the Wigner transform can be considered as the Fourier transform of $J(q,r)$. Hence, $W_\varphi(q,p)$ can also be expressed directly using the Fourier transform of the coordinate function F and it is then given by

$$W_\varphi(q,p) = \frac{1}{(2\pi)^3} \int_Q e^{iq\cdot v}\, \hat{F}\left(p - \frac{1}{2}v\right) \hat{F}\left(p + \frac{1}{2}v\right) dv. \qquad (3.182)$$

When the density $\varphi^\varepsilon = F^\varepsilon dq \in \mathrm{Den}(Q)$ depends on a scale parameter ε, the Wigner transform is defined as

$$W_\varphi^\varepsilon(q,p) = \frac{1}{(2\pi)^3} \int_Q e^{ip\cdot r}\, F^\varepsilon\left(q - \frac{\varepsilon}{2}r\right) F^\varepsilon\left(q + \frac{\varepsilon}{2}r\right) dr \qquad (3.183)$$

and also the scale dependence is lifted from configuration space to the cotangent bundle. The transform in Eq. 3.180 can be seen as a special case of Eq. 3.183 with $\varepsilon = 1$, and we will hence in the following only consider the scale dependent case.

The Wigner distribution $w_\varphi^\varepsilon \in \mathrm{Den}(T^*Q)$ of a configuration space density $\varphi^\varepsilon \in \mathrm{Den}(Q)$ is immediately related to observables which are quadratic in the density. For example, the "energy" or "position" density

$$n_\varphi^\varepsilon(q)\, dq = |F(q)|^2 \qquad (3.184)$$

is on phase space obtained from the fiber integral

$$n_\varphi^\varepsilon(q) = |F^\varepsilon(q)|^2 = \int_{T_q^*Q} w_\varphi^\varepsilon(q,p)\, dp. \qquad (3.185)$$

---

[201] The Wigner transform can also be defined with much weaker assumptions on $\star g$, see for example (Gérard et al., "Homogenization limits and Wigner transforms") and references therein, and it can clearly also be defined for functions.

Because of Eq. 3.185, the Wigner distribution $w_\varphi^\varepsilon \in \mathrm{Den}(T^*Q)$ is sometimes considered as a phase space representation of the "energy" density n_φ^ε, although it is important to note that it is not guaranteed to be positive. However, even negative values have a physical interpretation[202] and $w_\varphi^\varepsilon(q,p)$ is guaranteed to be positive as $\varepsilon \to 0$. When φ^ε is ε-oscillatory and its frequency is bounded satisfying

$$\overline{\lim_{\varepsilon \to 0}} \int_{\|p\| \geq R/\varepsilon} |\widehat{\varphi\, F^\varepsilon}|^2 dp \xrightarrow[R \to \infty]{} 0 \qquad (3.186)$$

for every continuous and compactly supported test function $\varphi \in \mathcal{F}(Q)$, so that the frequency decays sufficiently fast at infinity, then Eq. 3.185 carries over to the asymptotic limit and one has

$$n_\varphi^0(q) = \int_{T_q^*Q} w_\varphi^0(q,p)\, dp. \qquad (3.187)$$

Eq. 3.187 is a nontrivial result and was only recently established in the literature.[203] Analogous to the zero order moment of the Wigner distribution that gives the "energy" density, higher order moments yield additional observables that are quadratic in the field variables. For example, flux density $\vec{N}_\varphi^\varepsilon(q)$ is the first moment of the Wigner distribution and it is hence given by

$$\vec{N}_\varphi^\varepsilon(q) = \nabla_q \langle F^\varepsilon(q), F^\varepsilon(q) \rangle = \int_{T_q^*Q} \frac{p}{\|p\|}\, w_\varphi^\varepsilon(q,p)\, dp. \qquad (3.188)$$

The Wigner transform of a vector valued field

$$\Phi^\varepsilon(q,t) = (\Phi_1^\varepsilon(q,t), \cdots, \Phi_k^\varepsilon(q,t)) \in \mathcal{T}_0^k(Q) \qquad (3.189)$$

is defined by considering $\Phi^\varepsilon(q,t)$ component-wise with $\Phi_i^\varepsilon(q,t) \in \mathrm{Den}(Q)$.[204] The transform then yields a $k \times k$ Hermitian "matrix" density w_Φ^ε on phase space whose elements W_{ij}^ε are given by

$$W_{ij}^\varepsilon(q,p) = \frac{1}{(2\pi)^3} \int_Q e^{ip \cdot r}\, \Phi_i^\varepsilon\!\left(q - \frac{\varepsilon}{2}r\right) \Phi_j^\varepsilon\!\left(q + \frac{\varepsilon}{2}r\right) dr \in \mathrm{Den}(T^*Q), \quad (3.190)$$

---

[202] A now famous argument between Moyal and Dirac centered around the negative "probabilities" that exist in the phase space formulation of quantum mechanics, see the discussion in (Moyal, *Maverick Mathematician: The Life and Science of J.E. Moyal*, Chapter 3) and also (Feynman, "Negative probability").

[203] See (Tartar, "H-measures and Small Amplitude Homogenization"; Tartar, "H-measures, a New Approach for Studying Homogenization, Oscillations and Concentration Effects in Partial Differential Equations"; Gérard, "Microlocal defect measures"; Lions and Paul, "Sur les mesures de Wigner"; Gérard et al., "Homogenization limits and Wigner transforms") and references therein.

[204] For a rigorous discussion of the Wigner transform for non-scalar quantities see (Gérard, "Microlocal defect measures"; Lions and Paul, "Sur les mesures de Wigner"). A summary of the properties can be found in (Gérard et al., "Homogenization limits and Wigner transforms").

which is analogous to Eq. 3.180 with the integrand formed by two functions Φ_i^ε and Φ_j^ε instead of one.[205] Similar to the scalar case, W_Φ^ε becomes positive definitive only in the limit as $\varepsilon \to 0$. When the eigenvalues of $W_{ij}^\varepsilon(q,p)$ are simple, then the analogue of Eq. 3.185 for a vector valued field Φ^ϵ is given by

$$n_\Phi^\varepsilon(q,t) = \|\Phi^\varepsilon(q,t)\|^2 = \int \mathrm{tr}\left(W_\Phi^\varepsilon\right) dp \qquad (3.191a)$$

and it is again equivalent to the "energy" density $\|\Phi^\epsilon\|^2$. When the eigenvalue multiplicity is greater than one then

$$n_\Phi^\varepsilon(q,t) = \|\Phi^\varepsilon(q,t)\|^2 = \sum_a \int \mathrm{tr}\left(\Pi_a W_\Phi^\varepsilon\right) dp \qquad (3.191b)$$

where Π_a is the projection onto the eigenspace associated with the eigenvalue λ_a. As in the scalar case, the equality in Eq. 3.191 carries over to the asymptotic limit $\varepsilon \to 0$ when $\Phi^\varepsilon(q,t)$ is ε-oscillatory in each component.

---

[205]In the mathematics literature the scalar Wigner transform is often defined as a mapping $W(g,h) : \mathrm{Den}(Q) \times \mathrm{Den}(Q) \to \mathrm{Den}(T^*Q)$ which carries over more naturally to the vector-valued case but we prefer concreteness over generality for our treatment.

3.3 The Group Structure of Ideal Light Transport[206]

In the previous section, we derived a Hamiltonian formulation for light transport from Maxwell's equations, and we studied the geometric structure that arose in considerable detail. One of our results was the conservation of the light energy density—generalized radiance—along trajectories in phase space, see Fig. 3.10. By Noether's theorem, such a conserved quantity implies the existence of a symmetry, and the action of a Lie group on the phase space of the system. In the following, we will study the symmetry for light transport, which, as we will see, is the particle relabelling symmetry encountered elsewhere in continuum mechanics, and we will show that the group of symplectic transformations $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ of phase space is the symmetry group. When phase space is considered in its entirety, the group $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ also provides the configuration space for the system. Light transport is hence a Lie-Poisson system for the group $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ of symplectic transformations, and the associated Lie-Poisson momentum map provides a modern formulation of the constancy of light energy density along rays. A unitary representation for the action of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ on phase space T^*Q can be obtained using Stone's theorem, and for environments with surfaces the operator flow is defined locally. When combined with a scattering operator, this provides an operator theoretic description of light transport in such environments, paralleling those in the classical literature.

3.3.1 An Intuitive Justification

Monochromatic pre-radiance density, the five dimensional analogue of light energy density, is known as the light field or the plenoptic function, and it is well known that it provides a redundant description of light propagation since the light intensity is constant along a ray. Removing or exploiting the redundancy leads to a four dimensional description of the energy density, which in the literature has been called the photic field.[207] But redundancy implies symmetry, and hence light transport has to possess a symmetry group, and since light is constant along all trajectories, it has to be an infinite dimensional group.

To understand the symmetry group for light transport, one requires an alternative perspective on the configuration space of the system where the flow is considered in its entirety, and not only along each trajectory on the cotangent bundle T^*Q. When ℓ_0 is the light energy density at some initial time, then we know that its configuration at some later time t_1 is given by

$$\ell_0 = \eta_{t_1,0}^* \, \ell_1 \tag{3.192}$$

---

[206]The general theory of the material developed in this section was presented in Chapter 2.3.3 and Chapter 2.3.5, and the reader referred there for the necessary background.

[207](Moon and Spencer, *The Photic Field*); sometimes it is also referred to as the 4D light field (Levoy and Hanrahan, "Light Field Rendering").
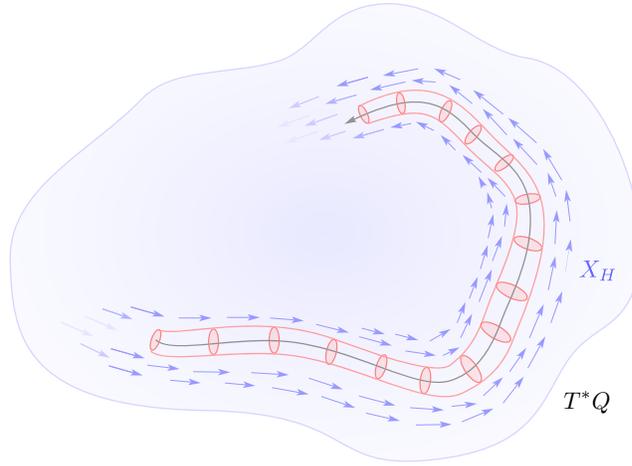
Figure 3.10: Conservation of the phase space energy density along the flow of the Hamiltonian vector field X_H.

where η_{t_1,t_0} is the flow from t_0 to t_1, see Fig. 3.4. Similarly, starting from $t = t_1$, we have for the light energy density ℓ_2 at some time t_2 that

$$\ell_1 = \eta_{t_2,t_1}^* \, \ell_2. \tag{3.193}$$

But, clearly, it is also satisfied that

$$\ell_0 = \eta_{t_2,0}^* \, \ell_{t_2}. \tag{3.194}$$

Hence, the pullback maps can be composed and we have for example

$$\ell_0 = \eta_{t_1,0}^* \, \ell_1 \tag{3.195a}$$

$$= \eta_{t_1,0}^* \, \eta_{t_2,t_1}^* \, \ell_2 \tag{3.195b}$$

$$= \eta_{t_2,t_1-t_1}^* \, \ell_2 \tag{3.195c}$$

$$= \eta_{t_2,0}^* \, \ell_2. \tag{3.195d}$$

and the combination of any two elements is again a valid map transporting light energy density. Since the maps η_{t_j,t_i} represent the flow of a time reversible Hamiltonian system, they have smooth inverses. But when we consider $\eta_{0,0}$ as the identity, then all the properties of a group—identity, composition, existence of inverse, closure—are satisfied, and since for every diffeomorphism there exists an infinitesimally far away diffeomorphism, which only differs slightly, it also makes intuitively sense that the diffeomorphisms form a smooth group, that is a Lie group. The diffeomorphisms of T^*Q we are considering are those generated by Hamiltonian vector fields, and the set of all such diffeomorphisms forms

the group $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ of symplectic transformations of phase space T^*Q.[208] Hence, with respect to ℓ_0, for all times t the configuration ℓ_t of the light energy density can be described by a diffeomorphism in $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$, and, globally considered, the group thus forms the configuration space for the system. Light transport is hence a Lie-Poisson system, and this implies that also the symmetry group for the system is $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$, which corresponds to the well known particle relabelling symmetry one encounters for example also in ideal fluid dynamics.

3.3.2   A Formal Proof

Following work for the Maxwell-Vlasov and Vlasov-Poisson system in plasma physics,[209] we will employ an indirect approach to establish that light transport is a Lie-Poisson system for the group $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ for symplectic diffeomorphisms: what we will show is that the light transport equation is the reduced Hamiltonian equation for the dynamics on the dual Lie algebra $\mathfrak{s}_+^*$ of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$, see again Fig. 2.49 for the intuition. Let us begin by formalizing the setting for ideal light transport.

Definition 3.1. *Ideal light transport is a Hamiltonian system satisfying:*

i) *The configuration space $Q \subset \mathbb{R}^3$ is a smooth, compact manifold with $\mathrm{H}^1(Q) = 1$, or $Q = \mathbb{R}^3$ with suitable decay conditions on the tensor fields, and the phase space is the usual cotangent bundle T^*Q with coordinates $(q, p) \in T^*Q$.*

ii) *The Hamiltonian is*

$$H(q, p) = \frac{c}{n(q)} \|p\|$$

   where the refractive index field $n : Q \to \mathbb{R}^+$ is a smooth field on Q.

iii) *The Hamiltonian vector field X_H is globally defined.*

For ideal light transport as defined in Def. 3.1, we consider the group $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ of symplectic diffeomorphisms generated by the Hamiltonian vector field X_H, and let the group act on the left on the cotangent bundle by $\eta \circ (q, p) = \eta(q, p)$. For the Lie algebra $\mathfrak{s}$ of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ we employ

$$\mathfrak{s} = (\mathcal{F}(T^*Q), \{\,,\}) \tag{3.196}$$

where $\{\,,\}$ is the canonical Poisson bracket on the cotangent bundle induced by the symplectic 2-from $\omega = \mathrm{d}\theta$, cf. Proposition 2.109. The dual Lie algebra is hence given by

$$\mathfrak{s}^* = \mathrm{Den}(T^*Q), \tag{3.197}$$

[208]We refer the reader again to Chapter 2.3.5.3 for the somewhat technical conditions when $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ is in fact the group of all diffeomorphisms generated by Hamiltonian vector fields.

[209]Marsden et al., "Hamiltonian Systems with Symmetry, Coadjoint Orbits and Plasma Physics".

cf. Proposition 2.110, and for the moment we will not distinguish between exact and arbitrary densities on phase space, and functions and densities on phase space will be identified when convenient, which is possible since the Liouville form ϖ is invariant under a Hamiltonian flow.

$\text{Diff}_{\text{can}}(T^*Q)$ is acting on the left on the cotangent bundle T^*Q, and hence the reduced dynamics are obtained in the right Lie algebra $\mathfrak{s}_+^*$, the Eulerian representation for the system, while the conserved quantities are defined in the left Lie algebra $\mathfrak{s}_-^*$, the convective representation, see again Fig. 2.50. Following the indirect approach we have chosen, we want to show that the reduced Hamiltonian equation in the Eulerian representation $\mathfrak{s}_+^*$ given by

$$\dot{\ell} = -\text{ad}_{\frac{\delta \mathcal{H}_\ell}{\delta \ell}}^*(\ell) = -\left\{\ell, \frac{\delta \mathcal{H}_\ell}{\delta \ell}\right\}, \tag{3.198}$$

cf. Proposition 2.111, coincides with the light transport equation in Eq. 3.39

$$\frac{\partial \ell}{\partial t} = -\left\{\ell, H\right\}. \tag{3.199}$$

Comparing Eq. 3.198 and Eq. 3.199 shows that we have to find a reduced Hamiltonian $\mathcal{H}_\ell : \mathfrak{s}_+^* \to \mathbb{R}$ on the dual Lie algebra $\mathfrak{s}^*$ whose derivate $\delta H_\ell/\delta \ell$ coincides with the Hamiltonian

$$H(q, p) = \frac{c}{n(q)}\|p\| = 2\pi\nu \tag{3.200}$$

in Eq. 3.38 that we employed so far for light transport. As we saw before, the "particle" Hamiltonian in Eq. 3.200 determines the trajectory for individual "particles" on phase space, but since it does not depend on the phase space light energy density it can clearly not be related to the total energy of ideal light transport as defined in Def. 3.1.

Without a Lagrangian formulation, there is no systematic approach to determine $\mathcal{H}_\ell$. However, with the previous results for light transport in mind, we propose—for the moment tentatively—that the Hamiltonian $\mathcal{H}_\ell$ is given by

$$\mathcal{H}_\ell = \mathcal{H}(\ell_t) = \int H(q, p)\, \ell_t(q, p)\, \varpi = \int \frac{c}{n(q)}\, \|p\|\, \ell_t(q, p)\, \varpi. \tag{3.201}$$

Indeed, since the particle Hamiltonian in Eq. 3.200 determines the energy for every phase space point (q, p) and we weight it by the density $\ell_t(q, p)$ which represents the amount of light at every point, Eq. 3.201 determines the total energy of ideal light transport, which provides a first validation for our "guess". We will refer to Eq. 3.201 in the following as the **field Hamiltonian for ideal light transport** since it related to the field formulation of the theory.

To obtain the reduced Hamiltonian equation in Eq. 3.198, we require the functional derivative $\delta\mathcal{H}_\ell/\delta \ell$. The usual ansatz yields

$$\left\langle \frac{\delta \mathcal{H}_\ell}{\delta \ell}, g \right\rangle = \left. \frac{d}{d\epsilon} \mathcal{H}_\ell(\ell + \epsilon g) \right|_{\epsilon = 0} \tag{3.202a}$$

$$= \frac{d}{d\epsilon} \int_{T^*Q} \frac{c}{n(q)} \, \|p\| \, (\ell + \epsilon g) \, \varpi \, \bigg|_{\epsilon=0} \tag{3.202b}$$

$$= \int_{T^*Q} \frac{c}{n(q)} \, \|p\| \frac{d}{d\epsilon} \, (\ell + \epsilon g) \, \varpi \, \bigg|_{\epsilon=0} \tag{3.202c}$$

$$= \int_{T^*Q} \frac{c}{n(q)} \, \|p\| \, g \, \varpi \tag{3.202d}$$

$$= \left\langle \frac{c}{n(q)} \, \|p\|, g \right\rangle \tag{3.202e}$$

and a comparison of the left hand side of Eq. 3.202a with Eq. 3.202e shows that

$$\frac{\delta \mathcal{H}_\ell}{\delta \ell} = H(q,p) = \frac{c}{n(q)} \, \|p\|. \tag{3.203}$$

As required, the functional derivative $\delta \mathcal{H}_\ell / \delta \ell$ hence coincides with the "particle" Hamiltonian H, which verifies our "guess". Inserting Eq. 3.203 in Eq. 3.198 then yields for the reduced Hamiltonian equation

$$\dot{\ell} = X_{\mathcal{H}_\ell}[\ell] = -\mathrm{ad}^*_{\frac{\delta \mathcal{H}_\ell}{\delta \ell}}(\ell) = - \left\{ \ell, \frac{\delta \mathcal{H}_\ell}{\delta \ell} \right\} = - \{\ell, H\} \tag{3.204}$$

and together with the general theory for $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ in Chapter 2.3.5.3 we established the following theorem.

Theorem 3.1 (Group Structure of Ideal Light Transport). *Ideal light transport is a Lie-Poisson system for the group* $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ *of symplectic transformations acting on the left on the cotangent bundle* T^*Q. *Time evolution is described by a curve*

$$\eta_t : [a, b] \to \mathrm{Diff}_{\mathrm{can}}(T^*Q)$$

and the light transport equation

$$\frac{\partial \ell}{\partial t} = - \{\ell, H\} \, .$$

describes the coadjoint action of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ *on the Eulerian representation* $\mathfrak{s}^*_+$. *Ideal light transport is right invariant under the action of* $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$.

An overview of the geometry of ideal light transport is hence provided in Fig. 2.50. With the dynamics defined in the Eulerian representation on $\mathfrak{s}^*_+$, missing is only the momentum map in the convective representation on $\mathfrak{s}^*_-$, and we will study it in the following section.

Remark 3.31. Our ansatz to establish the group structure for ideal light transport by showing that the light transport equation is a reduced Hamiltonian
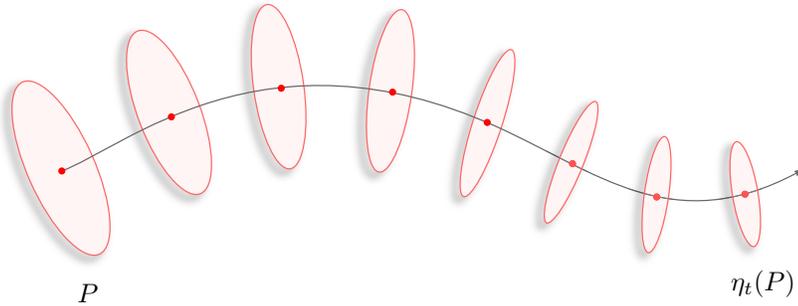
Figure 3.11: The momentum map for ideal light transport is equivalent to the constancy of the light energy density on a surface P under the Hamiltonian flow η_t, see also Fig. 3.10.

equation is analogous to how the group structure is derived for an ideal Euler fluid, where one shows that the ideal Euler fluid equation describes the coadjoint Hamiltonian flow

$$\omega_t = -\mathrm{ad}^*_{\frac{\delta H}{\delta \omega}} \omega = -\pounds_{v_t} \omega$$

where ω_t is the Eulerian vorticity and H the field Hamiltonian for fluid dynamics,[210] cf. Example 2.147.

3.3.3 The Momentum Map for Ideal Light Transport

Light transport is a Lie-Poisson system for the left action of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ and hence it is right invariant under the action of the group. By Noether's theorem, cf. Theorem 2.35, the momentum map $\mathbf{J}_R : T^*G \to \mathfrak{s}^*_-$ for the right action hence yields the conserved quantities, and these are defined in the convective representation $\mathfrak{s}^*_-$, Fig. 2.50. With $\mathfrak{s} = (\mathcal{F}(T^*Q), \{\,,\})$ and $\mathfrak{s}^* = \mathrm{Den}(T^*Q)$, the conserved quantity is the phase space light energy density $\ell^-_t \in \mathfrak{s}^*_-$ in the convective representation $\mathfrak{s}^*_-$. Moreover, since by the definition of the pullback,

$$(\eta_t^* \ell^-_t)(q,p) = \ell^-_t(\eta_t(q,p)), \tag{3.205}$$

the action of the map $\eta_t(q,p)$ can be considered as the push-forward of an arbitrary phase space volume whose points are (q,p), it is a modern formulation of the invariance of the phase space energy density along trajectories in phase space, cf. Fig. 3.11 The momentum map for light transport is hence also the direct analogue of Kelvin's circulation theorem, cf. Example 2.147.

Although the momentum map follows from the general theory, it is instructive to compute it also with "bare hands", similar to Example 2.147 where we explicitly computed the momentum map for ideal fluid dynamics. In contrast to the fluid case, however, for $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ defined on configuration space no

---

[210]Tronci, "Geometric dynamics of Vlasov kinetic theory and its moments", p. 33.

natural metric is available.[211] Hence, we have to work directly with the pairing $\langle\,,\rangle : \mathfrak{s} \times \mathfrak{s}^* \to \mathbb{R}$, which is well defined by the construction of the dual Lie algebra, cf. Chapter 2.3.5.3. By definition, the convective velocity for light transport is given by

$$\mathcal{X}_t = \eta_t^* X_t. \tag{3.206}$$

where X_t is the possibly time dependent Hamiltonian vector field corresponding to the Eulerian velocity, cf. Eq. 2.372. The light transport equation in the form of Eq. 3.50b in the convective representation is

$$0 = \eta_t^* \left(\frac{\partial \ell_t}{\partial t} + \pounds_{X_t} \ell_t \right) \tag{3.207a}$$

$$= \eta_t^* \left(\frac{d}{dt} \ell_t(\eta_t \bar{x}) \right) \tag{3.207b}$$

$$= \frac{d}{dt} \left(\eta_t^* \ell_t \right)(\bar{x}) \tag{3.207c}$$

$$= \frac{d}{dt} \mathrm{d}(\eta_t^* \alpha_t)(\bar{x}) \tag{3.207d}$$

and in the last line $\eta_t^* \alpha_t \in \Omega^{n-1}(T^*Q)$ is hence also a differential form in the convective representation, see Chapter 2.3.5.3 where we derived the dual Lie algebra $\mathfrak{s}^*$ for $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$. For the pairing between the left Lie algebra $\mathfrak{s}_-$ and its dual $\mathfrak{s}_-^+$ we hence have

$$\langle \mathcal{X}_t, \eta_t^* \alpha_t \rangle = \int_{T^*Q} i_{\mathcal{X}_t} \omega \wedge \eta_t^* \alpha_t \tag{3.208}$$

and since $i_{\mathcal{X}_t} \omega = i_{\eta_t^* X_t} \omega = \eta_t^* (i_{X_t} \omega)$, which holds by the invariance of the symplectic form ω under the action of η_t, we obtain

$$\int_{T^*Q} i_{\mathcal{X}_t} \omega \wedge \eta_t^* \alpha_t = \int_{T^*Q} \eta_t^* \left(i_{X_t} \omega \wedge \alpha_t \right). \tag{3.209}$$

With the change of variables theorem we have

$$\int_{T^*Q} i_{\mathcal{X}_t} \omega \wedge \eta_t^* \alpha_t = \int_{T^*Q} i_{X_t} \omega \wedge \alpha_t \tag{3.210}$$

with the right hand side being the pairing $\langle\,,\rangle : \mathfrak{s}_+ \times \mathfrak{s}_+^* \to \mathbb{R}$ in the Eulerian representation. Eq. 3.210 shows that

$$\mathcal{F}_t = \mathrm{d}(\eta_t^* \alpha_t) \tag{3.211}$$

---

[211]In Chapter 3.2.3 we introduced a natural metric g_n for light transport. However, it is defined on configuration space Q, and hence acts on tensors $\mathcal{T}_s^r(Q)$, such as vectors or 1-forms, on this space, but the Hamiltonian vector field is defined *on* T^*Q, that is $X_H \in T(T^*Q)$, where g_n is not available.
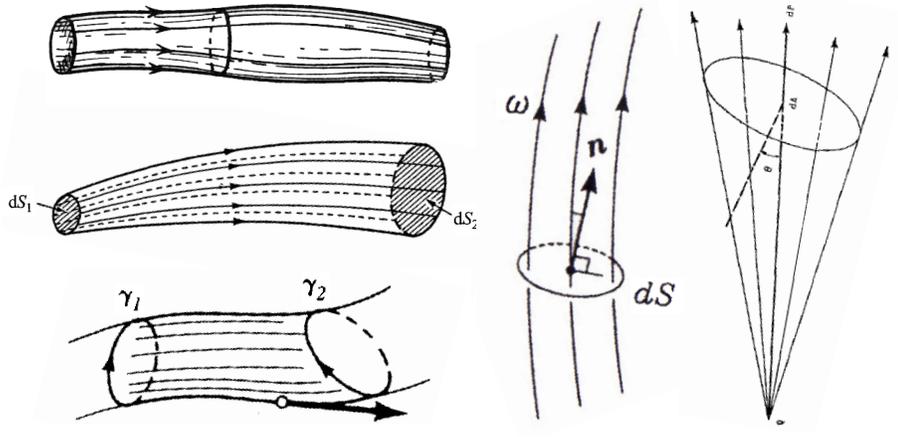
Figure 3.12: Classical renderings of fluid and light flow appear almost identical. Our results on the momentum map on light transport provide an explanation for this similarity.[‡]

is the convective phase space energy density that provides a non-degenerate pairing with the convective velocity $\mathcal{X}_t$. From Eq. 3.207d we therefore have

$$0 = \frac{d}{dt}\mathrm{d}(\eta_t^* \alpha_t)(\bar{x}) = \frac{d}{dt}\mathcal{F}_t \tag{3.212}$$

and $\mathcal{F}_t$ is conserved, which is the expected result from the general theory for Lie-Poisson systems. Analogous to Eq. 2.401 where we connected the momentum map for an ideal Euler fluid to Kelvin's circulation theorem, using the change of variables theorem for an arbitrary phase space P_0 in the convective or reference configuration yields

$$\int_{P_0} \mathcal{F}_t = \int_{P_0} \eta_t^* \ell_t = \int_{\eta_t(P_0)} \ell_t \tag{3.213}$$

which shows that the phase space light energy density ℓ_t is conserved under the transport $P_t = \eta_t(P_0)$ by the flow, cf. Fig. 3.11. A comparison between the situation for light and fluids is also shown in Fig. 3.12.

Remark 3.32. The light energy density is the momentum of light transport. In classical systems, such as a particle in a field which we considered before, the conservation of momentum is associated with translation invariance. A diffeomorphism group can be considered as a generalized, nonlinear translation invariance, where the nonlinearity leads to the more complex group picture.

[‡]Left top from (Milne-Thomson, *Theoretical Aerodynamics*, p. 168), left middle from (Born and Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, p. 123), left bottom from (Arnold, *Mathematical Methods of Classical Mechanics*, p. 237); middle from (Kambe, *Elementary Fluid Mechanics*, p. 84), right from (Nicodemus, "Radiance", p. 371).

With the momentum map for light transport we can consider the analogue of the enstrophy integrals in ideal fluid dynamics. In analogy to Eq. 2.402, a Casimir function $C_\phi(\ell_t)$ for ideal light transport is

$$C_\phi(\ell_t) = \int_{T^*Q} \phi\left(\frac{\ell_t}{\varpi}\right)^k \varpi \tag{3.214}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is an arbitrary smooth function, and the point-wise conservation follows from the change of variables theorem and since $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ acts via diffeomorphisms on its dual Lie algebra $\mathfrak{s}^*$.[212] For ϕ being a polynomial one obtains

$$C_k(\ell_t) = \int_{T^*Q} \ell_t^k \tag{3.215}$$

which corresponds to the statistical moments that have been considered in much previous works on kinetic theory.[213] Moreover, since the Casimirs are invariant under time evolution, with a finite number of the $C_k(\ell_t)$ one obtains a finite function space that is closed under the group action.

Remark 3.33. The existence of an analogue for enstrophy for light transport might suggest that also an analogue for helicity exists, which provide an important and interesting topological invariant for fluid flow. However, helicity only exists for three dimensional flow domains whereas phase space is always even dimensional.

3.3.4 An Operator Representation of Light Transport[214]

In the previous section, we established that $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ with its action on the cotangent bundle T^*Q is the configuration space for ideal light transport. In the following, we will obtain a representation of the action in a suitable Hilbert space $\mathcal{H}(T^*Q) \subset \mathcal{F}(T^*Q)$ which will provide an alternative, functional analytic description of the dynamics using linear operators acting on the space of light energy densities $\ell \in \mathcal{H}(T^*Q)$. We will obtain the representation by showing that the Hamiltonian vector field X_H for light transport can be interpreted as an anti-self-adjoint operator, which by Stone's theorem then implies that the flow map η_t corresponds to a one parameter group of real unitary operators $\mathcal{U}_t$ acting on $\mathcal{H}(T^*Q)$. Next to the additional insight which is obtained with the

---

[212]See (Khesin and Chekanov, "Invariants of the Euler equations for ideal or barotropic hydrodynamics and superconductivity in D dimensions") and (Arnold and Khesin, *Topological Methods in Hydrodynamics*, Proposition 9.3).

[213]See for example (Gibbons, "Collisionless Boltzmann Equations and Integrable Moment Equations"; Holm, Lysenko, and Scovel, "Moment invariants for the Vlasov equation"; Scovel and Weinstein, "Finite Dimensional Lie-Poisson Approximations to Vlasov-Poisson Equations"; Holm, Putkaradze, and Tronci, "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions") and references therein.

[214]The reader is referred to Chapter 2.3.2.7 for the necessary background for this section. As before, we will identify functions and densities when convenient.

functional analytic interpretation, the operator formalism enables to combine the Hamiltonian formulation of transport with a scattering operator describing the interaction at surfaces. This yields a description of light transport that includes scattering as a discrete time semi-group and it recovers the operator formulation known in the literature.

Stone's Theorem for Light Transport For Stone's theorem in Theorem 2.27 to apply we have to show that the infinitesimal generator A_H associated with the Hamiltonian vector field X_H is a anti-self-adjoint operator satisfying

$$\langle A_H f, g \rangle = \langle f, -A_H g \rangle. \tag{3.216}$$

Let $(\mathcal{H}(T^*Q), \langle , \rangle)$ be a Hilbert space with $\mathcal{H}(T^*Q) \subset (\mathcal{F}(T^*Q) \cap L_2(T^*Q))$ whose inner product $\langle , \rangle$ is the usual L_2 inner product. Then for $f, g \in \mathcal{H}(T^*G)$, the left hand side of Eq. 3.216 can by definition of the infinitesimal generator of a one-parameter semi-group in Def. 2.171 be written as

$$\langle A_H f, g \rangle = \langle X_H[f], g \rangle. \tag{3.217a}$$

With Proposition 2.94 we then obtain

$$\langle A_H f, g \rangle = \langle \{H, f\}, g \rangle \tag{3.217b}$$

and using the anti-symmetry of the Poisson bracket yields

$$\langle A_H f, g \rangle = \langle -\{f, H\}, g \rangle. \tag{3.217c}$$

By definition of the inner product, and with Corollary 2.21, we thus have

$$\langle A_H f, g \rangle = \int (-\{f, H\}, g)\, \mu \tag{3.217d}$$

$$= \int (-f\,\{H, g\})\, \mu. \tag{3.217e}$$

Using the above argument in the inverse order now gives

$$\langle A_H f, g \rangle = \langle f, -\{H, g\} \rangle \tag{3.217f}$$

$$= \langle f, -X_H(g) \rangle \tag{3.217g}$$

and we finally obtain

$$\langle A_H f, g \rangle = \langle f, -A_H g \rangle \tag{3.217h}$$

which shows that A_H is indeed anti-self-adjoint as required. We hence established the following proposition.

Proposition 3.1. *The representation of the Hamiltonian vector field X_H of ideal light transport on a Hilbert space $(\mathcal{H}(T^*Q), \langle , \rangle) \subseteq (\mathcal{F}(T^*Q) \cap L_2(T^*Q))$ is an anti-self-adjoint operator A_H.*

With Stone's theorem, a corollary of the above proposition is the following result which is of considerable importance for light transport.

Corollary 3.1. *Let $(\mathcal{H}(T^*Q), \langle , \rangle) \subseteq (\mathcal{F}(T^*Q) \cap L_2(T^*Q))$ be a Hilbert space and X_H the Hamiltonian vector field of ideal light transport whose representation on $\mathcal{H}(T^*Q)$ is the anti-self-adjoint operator A_H. Then time evolution of ideal light transport is represented by the **transport operator for light transport** or **light transport operator** which is an element in a one parameter group $\mathcal{U}(t) : \mathbb{R} \times \mathcal{H}(T^*Q) \to \mathcal{H}(T^*Q)$ of real unitary operators on $\mathcal{H}(T^*Q)$ given by*

$$\mathcal{U}_t = \mathcal{U}(t) = e^{tA_H} = \mathrm{id} + tA_H + \frac{1}{2!}(tA_H)^2 + \dots$$

and whose action is defined by

$$\mathcal{U}(t)\,\ell_t = \ell(\eta_t(z), t) = \eta_t^* \ell_t$$

for $\ell \in \mathcal{H}(T^*Q)$, and where $\eta_t : \mathbb{R} \times T^*Q \to T^*Q$ is the flow generated by X_H.

That A_H is a bounded operator, and hence $\mathcal{U}(t)$ can be represented using the exponential map, follows from its equivalence to $X_H[f]$ and the smoothness of the refractive index field $n(q)$ that defines the Hamiltonian vector field X_H. The definition of the action in the above corollary is due to Koopman.[215] An important consequence of the above corollary is that light transport is functional analytically closed, that is $\ell \in \mathcal{H}(T^*Q)$ remains in $\mathcal{H}(T^*Q)$ during transport by $\mathcal{U}_t$, cf Proposition 2.26.

Remark 3.34. A light transport operator $\mathcal{U}_t$ is a real unitary operator providing the time evolution map for the action of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ when ideal light transport is represented on a Hilbert space $\mathcal{H}(T^*Q)$. It can hence be considered as an infinite dimensional analogue of a rotation matrix for the rigid body, which provides a representation for the action of $\mathrm{SO}(3)$ on $\mathbb{R}^3$.

Light Transport as a Discrete Time Semi-Group One of the advantages of an operator theoretic formulation of light transport is that it enables to obtain a unified description of transport and scattering, which allows to describe light transport in environment with scattering surfaces as a discrete semi-group and recovers the operator formulation that is known in the literature. Let us begin by fixing the notion of an environment for light transport which we will require in this section.

Definition 3.2. *Let $Q \subset \mathbb{R}^3$ be a smooth, compact manifold, or $Q = \mathbb{R}^3$ with suitable decay conditions on tensor fields, and let $n \in \mathcal{F}(Q)$ be a smooth refractive index field. The **scene manifold** $(\mathcal{M}, \rho)$ with surface scattering kernel $\rho \in \Omega^5(T^-\mathcal{M}) \times \mathrm{Den}(T^*\mathcal{M})$ is the boundary ∂V of a possibly disconnected volume manifold $V \subset \mathbb{R}^3$, and the **light transport scene** or simply **scene** is the quadruple $(Q, n, \mathcal{M}, \rho)$.*

---

[215]See (Koopman, "Hamiltonian Systems and Transformations in Hilbert Space", p. 316); it is interesting to observe that Koopman begins his paper by considering the group of symplectic diffeomorphisms but most of his results are employed in the context of volume preserving diffeomorphisms.
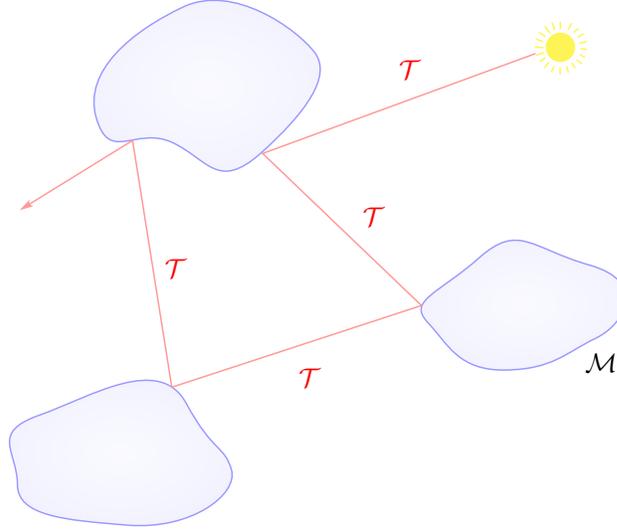
Figure 3.13: Light transport scene. The scattering light transport operator $\mathcal{T}$ is shown for a path in the scene, although it is important to keep in mind that it is acting on $\ell \in \mathrm{Den}(T^*Q)$, and hence on all paths simultaneously.

See Fig. 3.13 for the above definitions and Eq. 3.164 for the definition of a surface scattering kernel. We can now define a surface transport operator which describes light transport in between surfaces.

Definition 3.3. *Let* $(Q, n, \mathcal{M}, \rho)$ *be a light transport scene. Then the* **surface transport operator** $\bar{\mathcal{U}} : \mathcal{H}(T^+\mathcal{M}) \to \mathcal{H}(T^-\mathcal{M})$ *is*

$$(\bar{\mathcal{U}}f)(z) = (\mathcal{U}(t^{-1}(z))f)(z)$$

where $t^{-1}(z)$ *is the time from* $z \in T^-\mathcal{M}$ *to the previous surface, that is* $\eta_{t^{-1}(z)}(\bar{z}) = z$ *for some* $\tilde{z} \in T^+\mathcal{M}$.

In the definition, $T^+\mathcal{M}$ and $T^-\mathcal{M}$ denote again the half spaces of $T^*\mathcal{M}$ where $(\bar{p} \cdot \vec{n})$ is positive and negative, respectively, with $T^*\mathcal{M} = T^*Q|_{\mathcal{M}}$ being the restriction of T^*Q to the surface $\mathcal{M}$, and as usual we omitted inclusion maps $i_+ : \mathcal{H}(T^+\mathcal{M}) \to \mathcal{H}(T^*Q)$ and $i_- : \mathcal{H}(T^-\mathcal{M}) \to \mathcal{H}(T^*Q)$ that, strictly speaking, would have been necessary. It is important to note that the surface transport operator is no longer unitary. This results from its pointwise definition, and should be apparent when occlusion is considered, in which case the Hamiltonian vector field is no longer smooth. However, we have the following result for the surface transport operator.

Proposition 3.2. *Let* $(Q, n, \mathcal{M}, \rho)$ *be a light transport scene. Then the surface transport operator* $\bar{\mathcal{U}} : \mathcal{H}(T^+\mathcal{M}) \to \mathcal{H}(T^-\mathcal{M})$ *is an isometry, that is* $\|\bar{\mathcal{U}}\| = 1$.

The result follows from the conservation of the light energy density along trajectories in phase space.[216]

Next, we will introduce an operator formulation of scattering at surfaces, cf. Chapter 3.2.7.1, which, as we remarked before, does not admit a description using classical tools from analytic or geometric mechanics.

Definition 3.4. *Let $(Q, n, \mathcal{M}, \rho)$ be a light transport scene. Then the **surface scattering operator** $\mathcal{R}_\rho : \mathcal{H}(T^-\mathcal{M}) \to \mathcal{H}(T^+\mathcal{M})$ is*[217]

$$\hat{\ell} = \mathcal{R}_\rho \ell = \int_{T^+\mathcal{M}} \star\, (\mathrm{i}_{X_H} \ell) \wedge \rho.$$

It should be noted that the operator $\mathcal{R}_\rho$ defines scattering for all surface points $m \in \mathcal{M}$ simultaneously. We summarize the properties of the scattering operator in the following proposition.

Proposition 3.3. *Let $(Q, n, \mathcal{M}, \rho)$ be a light transport scene, and let $\mathcal{R}_\rho$ be the scattering operator for the scene and $\mathcal{H}(T^-\mathcal{M}) = L_2(T^-\mathcal{M})$. Then*

i) $\mathcal{R}_\rho$ is a Hilbert-Schmidt integral operator; that is, it is bounded and compact;

ii) $\mathcal{R}_\rho$ is self-adjoint;

iii) $\|\mathcal{R}_\rho\| < 1$.

The scattering operator $\mathcal{R}_\rho$ is Hilbert-Schmidt since the scattered light energy density has to be finite, self-adjointness follows from the requirements in Eq. 3.171 and i), and the finite norm follows from physical considerations, since no perfect scatterer exists and scattering cannot generate energy, see also Def. 2.40.[218]

Remark 3.35. For Proposition 3.3 we had to assume that $\mathcal{H}(T^-\mathcal{M}) = L_2(T^-\mathcal{M})$ for the scattering operator to be Hilbert-Schmidt, which is by definition an integral operator with symmetric kernel on $L_2(X)$ for some set X. For other spaces essentially the same results hold, see Remark 2.14.

We are now prepared to introduce an operator that completely describes light transport in a scene.

Definition 3.5. *Let $(Q, n, \mathcal{M}, \rho)$ be a light transport scene. The **scattering transport operator for light transport** $\mathcal{T} : \mathcal{H}(T^+\mathcal{M}) \to \mathcal{H}(T^+\mathcal{M})$ is*

$$\mathcal{T} = \mathcal{R}_\rho \bar{\mathcal{U}}.$$

---

[216]The result can also be found in (Arvo, "The Role of Functional Analysis in Global Illumination").

[217]We tacitly assume here that the space $\mathcal{H}$ is closed under scattering. This is a priori by no means guaranteed, but it might for example be possible to chose $\mathcal{H}$ large enough that it is satisfied.

[218]The properties in Proposition 3.3 were pointed out before by Arov ("The Role of Functional Analysis in Global Illumination").

By construction, when $\ell_0 \in \mathcal{H}(T^+\mathcal{M})$ is some initial phase space energy distribution, for example on light sources, then propagation is described by

$$\mathcal{T} \ldots \mathcal{T}\ell_0 = \mathcal{T}^k\ell_0 \tag{3.218}$$

and each application of $\mathcal{T}$ corresponds to one "bounce" in the scene, see again Fig. 3.13. Eq. 3.218 also defines an iterative map whose properties can be summarized as follows.

Proposition 3.4. *Let $(Q, n, \mathcal{M}, \rho)$ be a light transport scene with scattering transport operator $\mathcal{T}$. Then*

$$\mathcal{T}^i = \left\{ \begin{array}{ll} \mathrm{id} & i = 0 \\ \mathcal{T} \ldots \mathcal{T} & i > 0 \end{array} \right.$$

defines a discrete semi-group with composition $\mathcal{T}^i \circ \mathcal{T}^j = \mathcal{T}^{i+j}$.

That the definition of a discrete semi-group in Def. 2.171 is indeed satisfied is apparent. Because of the large value of the speed of light compared to macroscopic times, such as the time employed for physical measurements and the time over which ℓ_0 is constant, light propagation is almost instantaneous. Hence, at every location in a scene the phase space densities $\ell^k = \mathcal{T}^k\ell_0$ from different iterates are simultaneously present, and in fact this superposition is the observable in a light transport scene. This suggests the following definition.

Definition 3.6. *Let $(Q, n, \mathcal{M}, \rho)$ be a light transport scene with scattering transport operator $\mathcal{T}$. The **steady state phase space light energy density** $\bar{\ell} \in \mathcal{H}(T^+\mathcal{M})$ is*

$$\bar{\ell} = \ell_0 + \mathcal{T}^1\ell_0 + \mathcal{T}^2\ell_0 + \ldots = \sum_{i=0}^{\infty} \mathcal{T}^i\ell_0.$$

For the infinite sum on the right hand side in the above definition we have the following result.

Proposition 3.5. *Let $(Q, n, \mathcal{M}, \rho)$ be a light transport scene with scattering transport operator $\mathcal{T}$. Then the infinite series*

$$\sum_{i=0}^{n} \mathcal{T}^i \xrightarrow[n \to \infty]{} \mathcal{S}$$

*converges to the **solution operator** $\mathcal{S} : \mathcal{H}(T^+\mathcal{M}) \to \mathcal{H}(T^+\mathcal{M})$ in the usual operator topology.*

The norm of the scattering operator $\mathcal{R}_\rho$ is strictly bounded by unity by Proposition 3.3 and $\|\bar{\mathcal{U}}\| = 1$ by Proposition 3.2. Hence,

$$\|\mathcal{T}^i\| < \|\mathcal{T}^{i+1}\| \tag{3.219}$$

and Proposition 3.5 converges, at least formally, by the ratio test.[219] The definition of the steady state light energy density in Def. 3.6 can therefore also be written as

$$\bar{\ell} = \sum_{i=0}^{\infty} \mathcal{T}^i \ell_0 = \mathcal{S}\,\ell_0. \tag{3.220}$$

The above equation can be interpreted as a Neumann series, and it is hence equivalent to the following balance equation

$$(\mathrm{id} - \mathcal{T})\,\bar{\ell} = \ell_0 \tag{3.221a}$$

or more suggestively

$$\bar{\ell} = \ell_0 + \mathcal{T}\bar{\ell} \tag{3.221b}$$

which is the transport equation in operator form found in the classical literature.[220]

Remark 3.36. In Chapter 3.2.6 we defined measurements of the light energy density, but there we did not consider scattering and the resulting effect of a superposition of densities from different iterates of the scattering transport operator $\mathcal{T}$. As mentioned before, the observable in a light transport scene $(Q, n, \mathcal{M}, \rho)$ is hence the steady state light energy density $\bar{\ell}$ as defined in Def. 3.6 and Eq. 3.221.

3.3.5 Discussion

In this section we showed that ideal light transport is a Lie-Poisson system for the group $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ of symplectic diffeomorphisms on phase space T^*Q, and we established that the phase space energy density is the Noetherian quantity associated with the symmetry. To our knowledge, $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ has not been considered before as configuration space or symmetry group for light transport, neither in computer graphics nor in other fields where radiative transfer has been studied, and the conservation of the phase space energy density—generalized radiance—has in the past never been interpreted from the point of view of symmetries and conservation laws, that is from the point of view of modern physics. Our result establishes important parallels between light transport and other systems that also have a Lie-Poisson structure, and in particular systems in classical continuum mechanics such as the Maxwell-Vlasov and Vlasov-Poisson systems in plasma physics which also have $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ as symmetry group. Our derivations revealed a tantalizing similarity between the ideal Euler fluid and ideal light transport which suggests that the systems

---

[219]See for example (Smithies, *Integral Equations*, Chapter 2.5).

[220]In the computer graphics literature, the equation can already be found in Kajiya's landmark paper ("The Rendering Equation"), although it is well known for a long time in other fields.

can be considered as configuration and phase space analogues of each other.[221] For example, we showed that Kelvin's circulation theorem is equivalent to the conservation of the light energy density along trajectories in phase space, and both laws are corollaries of the right Lie-Poisson momentum map for the systems and describe the conservation of convective quantities. Despite the rich structure and parallels to other systems, it is currently not clear how valuable the group structure of ideal light transport will be, and one should rightfully be skeptical that it will provide the same theoretical and practical benefits that were obtained for fluid dynamics with the group perspective. Nonetheless, we believe that the structure has not been studied in sufficient detail to make a judgement, and many directions remain to be explored.

For example, in this section we established the Hamiltonian Lie-Poisson structure of ideal light transport. However, the Lagrangian formulation remains to be developed and it has to be studied if a Legendre transform connecting the two perspectives exists. Additionally, we believe that it will be possible to formulate ideal light transport using a semi-direct product $\text{Diff}_{\text{can}} \ltimes V^*$ where V^* is the space of phase space energy densities, cf. Def. 2.186, a structure which provides a very general model for continuum mechanics.[222]

One of the most intriguing features of the group structure of the ideal Euler fluid is that its time evolution is a geodesic flow on the group $\text{Diff}_\mu(Q)$ of volume preserving diffeomorphisms.[223] For light transport each trajectory on phase space is a geodesic with respect to the metric g_n defined by the refractive index, and one might hence similarly assume that time evolution for the system is a geodesic on $\text{Diff}_{\text{can}}(T^*Q)$. The main difficulty to establish such a result is that g_n is defined on configuration space Q while $\text{Diff}_{\text{can}}(T^*Q)$ defines a flow on phase space. Hence, the metric has to be lifted first to the cotangent bundle T^*Q before it can be employed for $\text{Diff}_{\text{can}}(T^*Q)$. Such a lift is known as Riemannian extension[224] and we consider it as a very interesting question for future work to investigate if light transport is a geodesic on $\text{Diff}_{\text{can}}(T^*Q)$ with respect to the lifted metric.[225] It has to be mentioned that on a Lie group a generalized notion of a geodesic exists which does not require a metric but which employs the connection defined by the translation action on the group.[226] By definition,

---

[221]In fact, in two dimensions the two groups are largely equivalent. A comparison between two dimensional Hamiltonian and volume preserving diffeomorphisms with a focus on global aspects can be found in (Gümral, "Geometry of plasma dynamics. I. Group of canonical diffeomorphisms").

[222]Holm, Marsden, and Ratiu, *The Euler-Poincaré Equations and Semidirect Products with Applications to Continuum Theories*.

[223]Systems whose time evolution is described by a geodesic on a group are sometimes referred to as Euler-Arnold systems (Tao, *The Euler-Arnold equation*). See (Khesin, "Topological Fluid Dynamics") for a list of known systems with this property.

[224]Patterson and Walker, "Riemann Extensions".

[225]Geodesics on Diff_{can} were recently also investigated by Ebin ("Geodesics on the Symplectomorphism Group") and Tronci ("Geometric dynamics of Vlasov kinetic theory and its moments") and co-workers.

[226]Rougly speaking, a connection on a manifold $\mathcal{M}$ is a structure that enables to define parallel transport on $\mathcal{M}$. On a Lie group, the left and right invariant vector fields provide a natural notion of parallel transport.

the exponential map on $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ defines a geodesic with respect to this connection. For a time invariant Hamiltonian vector field the flow is given by $\varphi_t = \exp(t\,X_H)$, and it can hence be considered as a Lie group geodesic, see also Remark 2.138. An important application of the group structure of the ideal Euler fluid have been results that establish the existence and uniqueness of solutions for the system.[227] It would be interesting to explore if the group structure of ideal light transport can yield similar insights.[228]

Following the literature, for the group structure of ideal light transport we employed the group $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ of symplectic diffeomorphisms and not, as would be more appropriate from a physical perspective, the group Γ introduced by van Hove, cf. Remark 2.184. The use of this group seems in particular important since we are working with $(\mathcal{F}(T^*Q), \{\,,\,\})$, which is more directly associated with Γ than with $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$. Similarly, in this section we considered $\mathfrak{s}^* = \mathrm{Den}(T^*Q)$ as dual Lie algebra for $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$, although we showed in Chapter 2.3.5.3 that the dual Lie algebra is formed only by exact volume forms $\Omega^n_{\mathrm{ex}}(T^*Q)$.[229] For fluid dynamics, the fact that the dual Lie algebra $\mathfrak{g}^*$ is formed by exact and not general 2-forms is critical in many contexts, cf. Example 2.147, and it leads for example to topological constraints on fluid flow.[230] It will be very interesting to explore if similar effects also arise for light transport, and an influence of the topology of a domain on the flow seems at least from a theoretical perspective very interesting. Another simplification we tacitly employed in the foregoing is the use of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$, despite it being "slightly too big" since our Hamiltonian is homogeneous of degree one and the Hamiltonian flow preserves the canonical 1-form θ and not only the symplectic 2-form ω. A more precise description would be obtained using the group $\mathrm{Diff}^{\theta}_{\mathrm{can}}(T^*Q \setminus \{0\}) \cong \mathrm{Diff}_{\mathrm{con}}(S^*Q)$ of contact diffeomorphisms of the cosphere bundle S^*Q.[231]

The group structure of light transport requires an idealized setting where the Hamiltonian vector field is globally defined. Although this is a very stringent requirement, for a compact domain $Q \subset \mathbb{R}^3$ it might be possible to physically realize such a vector field using so called meta-materials, that is materials whose

---

[227] Ebin and Marsden, "Groups of Diffeomorphisms and the Motion of an Incompressible Fluid".

[228] A classical proof for the existence and uniqueness of solutions for the transport problem expressed by the Vlasov equation can be found in (Dautray and Lions, *Mathematical Analysis and Numerical Methods for Science and Technology*, Chapter XXI). Ebin ("Geodesics on the Symplectomorphism Group") recently considered the existence and uniqueness problem for flows on $\mathrm{Diff}_{\mathrm{can}}(P)$.

[229] The difference between volume forms and densities is less of a concern in the present context. Strictly speaking, one would have to work with pseudo-forms, cf. Remark 2.87, but it is common to work with ordinary differential forms and interpret the result as volume form or density when integration is performed.

[230] For an overview of analogous questions for ideal fluid dynamics see (Khesin, "Topological Fluid Dynamics").

[231] Ratiu and Schmid studied the group structure for Hamiltonian systems homogeneous of degree one in their work on Fourier integral operators, see (Ratiu and Schmid, "The differentiable structure of three remarkable diffeomorphism groups") and also (Adams, Ratiu, and Schmid, "A Lie group structure for pseudodifferential operators"; Adams, Ratiu, and Schmid, *A Lie group structure for Fourier integral operators*).

refractive index takes values not found in nature,[232] and one would then obtain what could be called an "inverse cloaking device". In real-world environments, the Hamiltonian vector field is unfortunately never globally defined, since opaque objects discontinuously interrupt the flow. Nonetheless, from a practical point of view, many important aspects of the Lie-Poisson structure are still applicable when the diffeomorphisms are suitably restricted to flows in between surfaces, and for ideal fluid dynamics Kelvin's circulation theorem is a classical example although there also other work shows that results obtained in highly idealized environments can still provide many insights in less restrictive settings.[233] In the foregoing, we also developed a representation of the action of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ on phase space T^*Q using Hilbert space operators to adopt the idealized group structure to realistic environments which, when combined with the scattering operator, yielded a discrete semi-group given by the iterates of the scattering light transport operator. Our treatment of the semi-group structure was highly formal and requires considerable refinement, for example on what function spaces one has to consider. It should also be connected to the existing literature that employs semi-groups, often dissipative and contractive ones, to study transport theory.[234] To retain the group structure but nonetheless allow for effects such as scattering, it would be interesting to explore a formulation of light transport as a stochastic flow, cf. Remark 3.30.

We believe that the group structure of ideal light transport enables to develop structure preserving, numerical time integration algorithms, and similar to fluid dynamics the representation of the flow using a real, unitary operator might prove particularly useful in this respect.[235] To a limited extend the problem has been studied in the early 1990s for applications in plasma and astro-physics, based on general Lie-Poisson integrators developed at the time,[236] although to our knowledge only one of the approaches has been implemented.[237] An ansatz to develop structure preserving integrators is to employ moment expansions, which have been studied extensively in recent work,[238] and the resulting Lie algebra homomorphism from the Poisson bracket to the Schouten bracket. An interesting question for future work is thereby if bases other

---

[232]The concept of designing arbitrary refractive index fields emerged only very recently (Pendry, Schurig, and Smith, "Controlling Electromagnetic Fields"; Leonhardt, "Optical Conformal Mapping"); see also (Cui et al., *Metamaterials: theory, design, and applications*; Cai and Shalaev, *Optical Metamaterials*) for an up-to-date account of the state of the field.

[233]See for example (Mullen et al., "Energy-Preserving Integrators for Fluid Animation"; Pavlov et al., "Structure-preserving discretization of incompressible fluids").

[234]Engel and Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, Chapter VI.2.

[235]For ideal fluid dynamics, Koopman's work ("Hamiltonian Systems and Transformations in Hilbert Space") was employed to develop the only known structure preserving integrator (Mullen et al., "Energy-Preserving Integrators for Fluid Animation"; Pavlov et al., "Structure-preserving discretization of incompressible fluids").

[236]Zhong and Marsden, "Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators"; Channell and Scovel, "Integrators for Lie-Poisson dynamical systems".

[237]Channell, "Canonical Integration of the Collisionless Boltzmann Equation".

[238]See (Gibbons, Holm, and Tronci, "Geometry of Vlasov kinetic moments: A bosonic Fock space for the symmetric Schouten bracket"; Holm, Putkaradze, and Tronci, "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions") and references therein.

than polynomials can be employed, which should be possible since the Poisson bracket is bilinear, and if critical closure results[239] can be extended to such representations. An alternative research direction towards a discretization of the action of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ is to employ well known symplectic integrator that locally, for one trajectory in phase space, provide a structure preserving integrator. This could possible by extended to the action of $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ on a density, such as needed for light transport, with a reproducing kernel basis as introduced in the next chapter.[240] This would enable to reconstruct the density from the advected point samples at all times, with closure of the function space being guaranteed by existing results, cf. Theorem 2.17 and also our discussion following the generalized enstrophy integrals for light transport, and by Proposition 4.2 in the next chapter. A common ansatz to obtain discrete representations for fluid flow is to employ vortex sheets and filaments or point vortices. It seems that wave packets can be considered as an analogue to point vortices for light transport, and it would be interesting to explore if analogous to the fluid case special time evolution equations exist.

---

[239]Scovel and Weinstein, "Finite Dimensional Lie-Poisson Approximations to Vlasov-Poisson Equations"; Channell, "Canonical Integration of the Collisionless Boltzmann Equation".

[240]This was also suggested by Scovel and Weinstein ("Finite Dimensional Lie-Poisson Approximations to Vlasov-Poisson Equations") and Channell ("Canonical Integration of the Collisionless Boltzmann Equation") but these authors did not study the approach and they did not rigorously connect a particle approximation to a continuous phase space density, as is possible using reproducing kernel Hilbert spaces.

Chapter 4

Reproducing Kernel Bases for Light Transport Simulations

> *"John, when people thought the earth was flat,*
> *they were wrong. When people thought the earth was spherical,*
> *they were wrong. But if you think that thinking the earth*
> *is spherical is just as wrong as thinking the earth is flat,*
> *then your view is wronger than both of them put together."*[1]

In the last chapter, we derived light transport theory from Maxwell's equations, and we developed a modern, geometric formulation of the theory. In the following, we will reconsider the question of how light transport can be simulated numerically: how can we design effective computational techniques when only local information about the energy density is available? Our answer will be reproducing kernel bases, representations whose expansion coefficients are given by function values. Using such representations, we will establish finitary point functionals, constructive techniques that employ only local information, and which provide a common formulation for many classical algorithms, ranging from Monte Carlo integration over the Shannon sampling theorem to Gauss-Legendre quadrature. In contrast to the conceptualizations in the literature, our formulation enables to obtain close to optimal computational techniques by numerical optimization, and this allows for example to provide practical answers to the long standing questions on numerical integration in the box on the next page.[2] Moreover, performing Galerkin projection with reproducing kernel bases will provide a common functional analytic interpretation of a wide range of computational techniques for light transport simulation, including sampling based ones such as distribution ray tracing, path tracing, and photon mapping.

---

[1] Asimov, "The Relativity of Wrong".

[2] See (Dick and Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, p. 16) and (Shirley, Edwards, and Boulos, "Monte Carlo and Quasi-Monte Carlo Methods for Computer Graphics", p. 173).

1. How can we assess the quality of quadrature points?

2. How can we find quadrature points of particularly high quality?

3. Should samples on non-square domains be generated on the hypercube or on their native domains?

4. Should numeric optimization be used to generate samples?

5. Is there much to be gained from better sampling, or are we already in the diminishing return stage?

The emphasis in this chapter will be on constructive formulations, and we will hence forego with some of the mathematical rigour of the last chapter. A central question our treatment will thereby leave open—and where it very well might fall short of the reader's expectations—is that we do not characterize the function spaces we are working in and formally develop our arguments. Nonetheless, it is our belief that even this formal treatment provides vital scientific insight, and hence justifies—*temporarily*—the lacking rigour. Historically, such a mathematically questionable approach seems in fact often to be unavoidable to make progress on a subject, and well known examples for groundbreaking work where functional analytic questions had to be left open initially are Feynman path integrals in quantum mechanics,[3] applications of infinite dimensional Lie groups in geometric mechanics,[4] and the transition to geometric optics in electromagnetic theory.[5]

We will begin this chapter by introducing reproducing kernel bases and by developing finitary point functionals in Chapter 4.2. There, we will also present two practical applications where we employ the theory for numerical computa-

---

[3]See for example (Albeverio and Mazzucchi, *Path integral: mathematical aspects*) for a discussion of the functional analytic questions still open in the context of path integrals.

[4]When Arnold introduced the idea to employ diffeomorphism groups to describe dynamical systems, he formulated the situation as follows: *"Bien que* SDiff(D) *ne soit un groupe de Lie, il est intéressant de voir, en quelles affirmations se transforment les formules des paragraphes précédent dans ce cas. Une telle étude «euristique» amène aux propositions qu'on peut vérifier ensuite rigoureusement [. . .]."*, (Arnold, "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits", p. 340). Many of the functional analytic questions were settled shortly afterwards in a landmark paper by Ebin and Marsden ("Groups of Diffeomorphisms and the Motion of an Incompressible Fluid"), see also our discussion in Chapter 2.3.3.3, although it is fair to say that a general theory is still open.

[5]Only in the 1980s, in the work on *H*-measures and microlocal defect measures that we discussed before in Chapter 3, and more than 75 years after the original work on the subject by Sommerfeld and Runge ("Anwendung der Vektorrechnung auf die Grundlagen der geometrischen Optik"), has a derivation of the eikonal equation from electromagnetic theory been established which does not rely on unrealistic assumptions on the smoothness of the functions involved, cf. (Tartar, "Mathematical Tools for Studying Oscillations and Concentrations: from Young Measures to H-Measures and Their Variants"; Tartar, *The General Theory of Homogenization: a personalized introduction*).

tions. Subsequently, in Chapter 4.3, we will develop our unified formulation of computational techniques for light transport. The continuous formulation from which we will thus depart is the scattering transport operator that describes light propagation as a "flow" on the infinite dimensional space of light energy densities in a scene with scattering surfaces. As in previous work, the transition to a finite description is then possible using Galerkin projection and by representing the energy density as a basis expansion. Where we will differ from the literature, is the use of reproducing kernel bases where the expansion coefficients are given by function values. Algorithms such as distribution ray tracing, path tracing, and photon mapping then admit a functional analytic interpretation, and Galerkin projection can be employed for a mathematically and conceptually consistent treatment of computational techniques for light transport simulation. Our current, albeit limited, understanding of the function spaces of light transport will be presented in Chapter 4.4. We conclude the chapter with a discussion of the presented material in Chapter 4.5.

4.1 Reproducing Kernel Bases[6]

In this section, we will develop and study the central pillar of the work in this chapter: biorthogonal and possibly overcomplete bases formed by reproducing kernel functions. In the following, $\mathcal{H}(X)$ will be a reproducing kernel Hilbert space defined over the set X with dimension n, where n is possibly infinity, for which the reproducing kernel is given by $k_y(x) = k(y, x)$.

4.1.1 Construction

We are interested in the representation of arbitrary functions $f \in \mathcal{H}(X)$ using the reproducing kernel $k_y(x)$ for the space. Assuming for the moment that $\mathcal{H}(X)$ is finite dimensional, then we require n linearly independent functions for such a representation. But n functions in $\mathcal{H}(X)$ are easily obtained from the kernel functions $k_y(x)$ by choosing a set $\Lambda = \{\lambda_i\}_{i=1}^n$ of n points $\lambda_i \in X$, and then "anchoring" the reproducing kernels $k_i(x) = k_{\lambda_i}(x)$ at these locations. Assuming the so defined functions are linearly independent, we can represent any $f \in \mathcal{H}(X)$ using the set $\{k_{\lambda_i}(x)\}_{i=1}^n$. The utility of this construction will become apparent shortly in Proposition 4.1. Before, however, we will formalize the notion of a basis formed by reproducing kernel functions.

Definition 4.1. *Let $\mathcal{H}(X)$ be a reproducing kernel Hilbert space defined over the set X with reproducing kernel $k_y(x) = k(y, x)$. A **reproducing kernel basis** over a countable set $\Lambda = \{\lambda_i\}_{i=1}^m$ of m **reproducing points** $\lambda_i \in X$ is a frame for $\mathcal{H}(X)$ with basis functions*

$$k_i(x) = k(\lambda_i, x).$$

---

[6] See Chapter 2.2.3 for an introduction to reproducing kernel Hilbert spaces. Following the functional analysis and signal processing literature, we will here and in the sequel not distinguish between contravariant and covariant objects, and the location of an index will be chosen based on convenience.

For notational convenience, and motivated by the close analogy between bases and frames developed in Chapter 2.2.2.4, the above definition employs 'basis' even when $m > n$ and the kernel functions $k_i(x) = k(\lambda_i, x)$ provide an overcomplete representation. From our discussion on Riesz bases and frames in Chapter 2.2.2.3 and Chapter 2.2.2.4, we have the following, central result.

Proposition 4.1. *Let $\{k_i(x)\}_{i=1}^m$ be a reproducing kernel basis over the reproducing kernel Hilbert space $\mathcal{H}(X)$. Then there exist **dual reproducing kernel functions** $\tilde{k}_i(x) \in \mathcal{H}(X)$ such that every function $f \in \mathcal{H}(X)$ admits a representation*

$$f(x) = \sum_{i=1}^m \langle f(y), k_i(y) \rangle \, \tilde{k}_i(x) = \sum_{i=1}^m f(\lambda_i) \, \tilde{k}_i(x).$$

The basis expansion in the above proposition is at the heart of the present chapter: rather than computing inner products to obtain basis function coefficients, with a reproducing kernel bases these are given by the values $f(\lambda_i)$ of the function at the reproducing points λ_i. In the following, with slight abuse of notation, we will often refer to the basis pair formed by $\{k_i(x)\}_{i=1}^m$ and $\{\tilde{k}_i(x)\}_{i=1}^m$ as a reproducing kernel basis. A corollary of Proposition 4.1 is the following.

Corollary 4.1. *Let $(k_i(x), \tilde{k}_i(x))_{i=1}^n$ be a reproducing kernel basis which is a Riesz basis for the Hilbert space $\mathcal{H}(X)$. Then the dual kernel functions $\tilde{k}_i(x)$ are defined by*

$$\langle k_i(x), \tilde{k}_j(x) \rangle = \delta_{ij}$$

and they are interpolatory satisfying

$$\tilde{k}_j(\lambda_i) = \delta_{ij}.$$

The second part of the above corollary follows from the reproducing property of the $k_i(x)$,

$$\delta_{ij} = \langle k_i(x), \tilde{k}_j(x) \rangle = \tilde{k}_j(\lambda_i).$$

Unfortunately, the interpolation property does not hold for an overcomplete representation since biorthogonality is then no longer satisfied.

In analogy to Proposition 2.26, which states that a unitary operator preserves the properties of a Hilbert space frame, for reproducing kernel bases we have the following result for a unitary operator that arises from a pointwise action on the underlying set.

Proposition 4.2. *Let $\mathcal{H}(X)$ be a reproducing kernel Hilbert space defined over the set X, and let $\{k_i(x)\}_{i=1}^m$ be a reproducing kernel basis for the space with frame bounds A, B. Moreover, let the one parameter group of real unitary operators $U_t : \mathbb{R} \times \mathcal{H}(X) \to \mathcal{H}(X)$ be generated by a flow $\varphi_t : \mathbb{R} \times X \to X$ by*

$$(U_t f)(x) = f(\varphi_t(x)) = \varphi_t^* f(x).$$
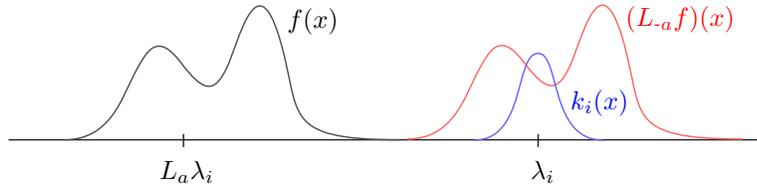
Figure 4.1: Action of the unitary left translation operator $L_a : \mathcal{H}(\mathbb{R}) \to \mathcal{H}(\mathbb{R})$ and effect on a reproducing kernel function $k_i(x) = k_{\lambda_i}(x)$.

Then the sequence $\{U_t\, k_i(x)\}_{i=1}^m$ generated by U_t, for arbitrary t, is again a reproducing kernel basis for $\mathcal{H}(X)$ with frame bounds A, B.

Proof. By Proposition 2.26, we know that U_t preserves the frame bounds of $\{k_i(z)\}_{i=1}^m$. For an arbitrary function $f \in \mathcal{H}(X)$, we have by Proposition 2.15 also that

$$\langle f(x), \mathcal{U}_t\, k_i(x) \rangle = \langle \mathcal{U}_{-t} f(x),\, k_i(x) \rangle. \tag{4.1a}$$

By the definition of $\mathcal{U}_t$ this is equivalent to

$$\langle f(x), \mathcal{U}_t\, k_i(x) \rangle = \langle \varphi_{-t}^* f(x),\, k_i(x) \rangle \tag{4.1b}$$

and with the definition of the pullback for functions we obtain

$$\langle f(x), \mathcal{U}_t\, k_i(x) \rangle = \langle f(\varphi_{-t} x),\, k_i(x) \rangle. \tag{4.1c}$$

Through the inverse shift by η_{-t}, cf. Fig. 4.1, one therefore has

$$\langle f(x), \mathcal{U}_t\, k_i(x) \rangle = f(\varphi_t \lambda_i) \tag{4.1d}$$

which shows that the flow by $\mathcal{U}_t$ also preserves the reproducing property of the functions $k_i(x)$, and these are then "anchored" at the time evolved reproducing points $\varphi_t(\lambda_i)$. $\qquad\blacksquare$

Proposition 4.2 has applications for example when the flow $\varphi_t : \mathbb{R} \times X \to X$ is a one parameter group of diffeomorphisms on a manifold and U_t is the functional analytic description of the flow obtained using "Koopmanism", cf. Chapter 2.3.2.7. For numerical computations, the proposition enables to compute the time evolution of a continuous function by time integrating the reproducing points λ_i instead of explicitly discretizing the unitary operator U_t.

From Def. 4.1 and Proposition 4.1, it is apparent that even when the Hilbert space $\mathcal{H}(X)$ is fixed, the properties of a reproducing kernel basis—and the very question of existence of such a basis—depends critically on the set $\Lambda \subset X$ of reproducing points that is employed. In the spirit of the constructive programme of this section, we will relegate the construction of reproducing points to Chapter 4.1.3 and general existence questions to Chapter 4.2.4, and instead continue by developing how reproducing kernel bases can be employed for computations.

4.1.2 Reproducing and Reconstruction Matrices

To understand how numerical computations can be performed with reproducing kernel bases, we will in the following adapt the concepts developed in Chapter 2.2.4 to these representations. For this section, let $\mathcal{H}(X)$ be a finite dimensional Hilbert space of dimension n, and let $\{\phi_i\}_{i=1}^n$ be an orthonormal basis for $\mathcal{H}(X)$.

By Proposition 2.31, every finite dimensional Hilbert space is a reproducing kernel Hilbert space, and by Proposition 2.36, the reproducing kernel is in this case given by

$$k_y(x) = k(y, x) = \sum_{i=1}^{n} \phi_i(y)\, \phi_i(x). \tag{4.2}$$

Moreover, with our results from Chapter 2.2.4.2, numerical computations in $\mathcal{H}(X)$ can be performed using the isomorphism between the space and $\mathbb{R}^n$ that is established by $\{\phi_i\}_{i=1}^n$. On $\mathbb{R}^n$, a representation of a reproducing kernel basis $\{k_i(x)\}_{i=1}^m$ is thereby given by the basis matrix from Def. 2.39, and since the $\phi_i(y)$ in Eq. 4.2 can be considered as the expansion coefficients of the reproducing kernel $k_y(x)$ with respect to the basis functions $\phi_i(x)$, cf. Remark 2.28, the elements of the matrix are $\phi_i(\lambda_j)$. We summarize these observations in the following definition.

Definition 4.2. *Let $\mathcal{H}(X)$ be a finite dimensional Hilbert space of dimension n, and let $\{\phi_i\}_{i=1}^n$ be an orthonormal basis for $\mathcal{H}(X)$. For a reproducing kernel basis $\{k_i(x)\}_{i=1}^m$ for $\mathcal{H}(X)$ defined over the points $\Lambda = \{\lambda_i\}_{i=1}^m$, the **reproducing matrix** $\mathrm{K}_\phi(\Lambda)$ is*

$$\mathrm{K} = \mathrm{K}_\phi(\Lambda) = \begin{bmatrix} \phi_1(\lambda_1) & \cdots & \phi_n(\lambda_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\lambda_m) & \cdots & \phi_n(\lambda_m) \end{bmatrix} \in \mathbb{R}^{m \times n}$$

and it provides the change of basis from $\{\phi_i\}_{i=1}^n$ to $\{k_i(x)\}_{i=1}^m$.

By definition of the basis matrix, the reproducing matrix satisfies

$$\mathrm{f}(\Lambda) = \mathrm{K}_\phi(\Lambda)\,\mathrm{f}(\phi) \tag{4.3}$$

where $\mathrm{f}(\phi)$ and $\mathrm{f}(\Lambda)$ are the coefficient vectors with respect to the bases, cf. Chapter 2.2.4.2, and we write $\mathrm{f}(\Lambda)$ since the coefficients of the kernel basis are the function values at the locations $\lambda_i \in \Lambda$. In fact, writing Eq. 4.3 in index notation yields

$$f_i(\Lambda) = \sum_{j=1}^{n} k_{ij} f_j(\phi) \tag{4.4a}$$

$$= \sum_{j=1}^{n} \phi_j(\lambda_i)\, f_j(\phi) \tag{4.4b}$$

$$= \sum_{j=1}^{n} f_j(\phi)\, \phi_j(\lambda_i) \tag{4.4c}$$

which is just the reconstruction formula for $f \in \mathcal{H}(X)$ at the sampling locations λ_i, and this are indeed the basis function coefficients $f_i(\Lambda) = f(\lambda_i)$. Eq. 4.4 also justifies our nomenclature for $\mathrm{K}_\phi(\Lambda)$ in Def. 4.2, since the matrix indeed reproduces the function values $f(\lambda_i)$ from the basis function coefficients $f_j(\phi)$.

Remark 4.1. In contrast to a general basis matrix $B(\psi, \phi)$ where the elements are given by $b_{ij} = \langle \psi_i, \phi_j \rangle$, the reproducing matrix can be determined by evaluating the function $\phi_i(\lambda_j)$ at the location λ_j and without the computation of inner products. This is analogous to Proposition 4.1, and that the reproducing matrix can be constructed easily and without approximations is vital for the practicality of reproducing kernel bases.

As will be shown shortly in Corollary 4.2, the inverse of the kernel matrix plays a special role and it hence deserves a name of its own.

Proposition 4.3. *Let $\mathcal{H}(X)$ be a finite dimensional reproducing kernel Hilbert space of dimension n with orthonormal basis $\{\phi_i\}_{i=1}^{n}$, and let $\mathrm{K}_\phi(\Lambda)$ be the reproducing matrix for a reproducing kernel basis $\{k_i\}_{i=1}^{m}$. The **reconstruction matrix** $\mathrm{R}_\phi(\Lambda)$ is the inverse of the kernel matrix*

$$\mathrm{R}_\phi(\Lambda) = \mathrm{K}_\phi^{-1}(\Lambda) \in \mathbb{R}^{n \times m}$$

and it provides the change of basis from $\{k_i\}_{i=1}^{m}$ to $\{\phi_i\}_{i=1}^{n}$. Moreover, the columns of $\mathrm{R}_\phi(\Lambda)$ are the basis function coefficients

$$r_{ij} = \langle \tilde{k}_i, \phi_j \rangle$$

of the dual reproducing kernel functions $\tilde{k}_i$ with respect to $\{\phi_i\}_{i=1}^{n}$.

The proposition follows immediately from Def. 2.39 in Chapter 2.2.4.2. The crucial consequence of the above result for computations is that

$$\mathrm{f}(\phi) = \mathrm{K}_\phi^{-1}(\Lambda)\, \mathrm{f}(\Lambda) = \mathrm{R}_\phi(\Lambda)\, \mathrm{f}(\Lambda) \tag{4.5}$$

and the basis function coefficients $f_i(\phi)$ of a function $f \in \mathcal{H}(X)$ with respect to an orthonormal basis $\{\phi_i\}_{i=1}^{n}$ can be obtained from the values of $f(\lambda_i)$ at the reproducing points λ_i, values which are much more easily obtained in practice than inner products. This observation is important enough to be stated as a result in its own right.

Corollary 4.2. *Let $\mathcal{H}(X)$ be a finite dimensional reproducing kernel Hilbert space of dimension n with orthonormal basis $\{\phi_i\}_{i=1}^{n}$, and let $\mathrm{K}_\phi(\Lambda)$ be the kernel matrix for a reproducing kernel basis $\{k_i(x)\}_{i=1}^{m}$. Then for $f \in \mathcal{H}(X)$, the basis function coefficients with respect to $\{\phi_i\}_{i=1}^{k}$ are given by*

$$f_i(\phi) = \sum_{j=1}^{m} r_{ij}\, f(\lambda_j)$$

where $\mathrm{R}_\phi(\Lambda) = \mathrm{K}_\phi^{-1}(\Lambda)$ is the reconstruction matrix with elements r_{ij}.

The above result can be extended to arbitrary, biorthogonal bases $\{\psi_i\}_{i=1}^n$ for $\mathcal{H}(X)$ when the reproducing kernel is expanded as in Remark 2.26. Note also that while n function values $f(\lambda_i)$ are sufficient, we can employ more when the reproducing kernel basis is a frame. We will see in the following that this is often advantageous, and the situations where this is the case are closely related to the robustness which motivates the use of frames in other applications.

Remark 4.2. As discussed before in Remark 2.33, when $m > n$ and the kernel basis is a frame, then the reconstruction matrix has to be determined using a pseudo-inverse.

Remark 4.3. The constructions in this section do not require that the Hilbert space is finite, and a separable reproducing kernel Hilbert space would have sufficed.[7] The infinite matrices one obtains in this case are no longer useful for numerical computations, but they are beneficial for example because their analysis is often easier than those of abstract operators.

4.1.3   Tightness of Reproducing Kernel Bases

The existence of a reproducing kernel basis rests on a set $\Lambda = \{\lambda_i\}$ of locations $\lambda_i \in X$ such that the $k_i(x) = k(\lambda_i, x)$ form a frame for the space $\mathcal{H}(X)$. For an arbitrary space $\mathcal{H}(X)$, it will not be possible to employ any set of locations, and we will discuss the general situation to a certain extent in Chapter 4.2.4. In this section, we will again concentrate on the finite dimensional case and on constructive results. Let us begin by introducing a name for point sets for which the reproducing kernel functions form a frame for $\mathcal{H}(X)$.

Definition 4.3. *Let $\mathcal{H}(X)$ be a reproducing kernel Hilbert space defined over the set X. When the countable set $\Lambda = \{\lambda_i\}$ of locations $\lambda_i \in X$ yields a reproducing kernel basis for $\mathcal{H}(X)$, then Λ forms a **spanning point set for** $\mathcal{H}(X)$.*

Using the reproducing matrix $\mathrm{K}(\Lambda)$ introduced in the last section, a simple criterion for a point set to be a spanning point set is the following.

Proposition 4.4. *Let $\mathcal{H}(X)$ be a finite dimensional reproducing kernel Hilbert space of dimension n defined over the set X, and let $\{\phi_i\}_{i=1}^n$ be an orthonormal basis for $\mathcal{H}(X)$. Then a set $\Lambda = \{\lambda_i\}$ of locations $\lambda_i \in X$ is a spanning point set for $\mathcal{H}(X)$ if and only if the determinant $\det(\mathrm{K}_\phi(\Lambda))$ of the reproducing matrix satisfies*

$$\det(\mathrm{K}_\phi(\Lambda)) \neq 0.$$

The reconstruction of basis function coefficients $f_j(\phi)$ from function values $f(\lambda_i)$ in Corollary 4.2 is equivalent to the solution of the linear system

$$f(\Lambda) = \mathrm{K}_\phi(\Lambda)\, f(\phi).$$

---

[7]Cf. (Meschkowski, *Hilbertsche Räume mit Kernfunktion*).

and this immediately establishes the above proposition since the definition of a frame asks that such a reconstruction is possible—this was the intrinsic meaning of the frame bounds.

Numerically and practically, not only is the existence of a solution to a linear system of relevance, but so too its "quality". This suggests to employ the condition number $\text{cond}(K_\phi(\Lambda))$ of the reproducing matrix $K_\phi(\Lambda)$, which determines how accurately the linear system can be solved, so as to distinguish between "good" and "bad" sequences of locations λ_i.

Remark 4.4. The condition number $\text{cond}(A)$ for the solution of a linear system represented by a matrix A can be defined as

$$\text{cond}(A) = \frac{\sigma_n}{\sigma_1}$$

where σ_n and σ_1 are the largest and smallest singular values of A, respectively, or, equivalently, the eigenvalues of the Gramian matrix $G = A^T A$. By Proposition 2.6, $\sigma_1 \neq 0$ if and only if A is invertible and the kernel $\ker(A)$ of the matrix A is trivial. Hence, an alternative characterization of spanning point sets is

$$\text{cond}(K_\phi(\Lambda)) \neq 0.$$

Using the condition number as quality measure, we can define optimal spanning point sets which define tight frames, cf. Chapter 2.2.2.4.

Proposition 4.5. *Let $\mathcal{H}(X)$ be a finite dimensional reproducing kernel Hilbert space defined over the set X, and let $\{\phi_i\}_{i=1}^n$ be an orthonormal basis for $\mathcal{H}(X)$. A set of locations $\Lambda = \{\lambda_i\}_{i=1}^m$ in X is an **extremal spanning point set** $\hat{\Lambda}$ when the kernel matrix $K_\phi(\hat{\Lambda})$ satisfies*

$$\text{cond}(K_\phi(\hat{\Lambda})) = 1$$

*and all eigenvalues λ_i of $K_\phi(\hat{\Lambda})$ are equal. Moreover, the reproducing kernel basis $\{k_i(x)\}_{i=1}^m$ defined by $\hat{\Lambda}$ then forms a tight frame, and when the **tight reproducing kernel basis** over $\hat{\Lambda}$ is normalized such that $\|k_i(x)\| = 1$, then $\lambda_i = m/n$.*

The proposition follows immediately from the literature,[8] and it is interesting to note that extremal spanning point sets are also minimizers of the frame potential, cf. Remark 2.23.[9]

---

[8](Benedetto and Fickus, "Finite Normalized Tight Frames", p. 375) and (Sloan and Womersley, "Extremal Systems of Points and Numerical Integration on the Sphere", p. 5).

[9]This follows from (Benedetto and Fickus, "Finite Normalized Tight Frames", Theorem 7.1).

Remark 4.5. For an extremal point set the inverse kernel matrix is given by the transpose $K_\phi^T(\Lambda)$, that is

$$K_\phi^{-1}(\Lambda) = K_\phi^T(\Lambda) = \begin{bmatrix} \phi_1(\lambda_1) & \cdots & \phi_1(\lambda_m) \\ \vdots & \ddots & \vdots \\ \phi_n(\lambda_1) & \cdots & \phi_n(\lambda_m) \end{bmatrix} \in \mathbb{R}^{n \times m}$$

and hence by Corollary 4.2 the basis function coefficients are obtained by

$$f_i(\phi) = \sum_{j=1}^m r_{ij} \, f(\lambda_j) = \sum_{j=1}^m \phi_i(\lambda_j) \, f(\lambda_j)$$

which is particularly easy to implement numerically and does not require the solution of a linear system.

Similar to orthogonal bases, tight reproducing kernel frames are difficult to obtain and do not allow for other desirable properties. Hence, in practice usually reproducing kernel bases defined over sampling sequences which yield a condition number close to unity are employed. This motivates the following definition.

Definition 4.4. *Let $\mathcal{H}(X)$ be a finite dimensional reproducing kernel Hilbert space of dimension n defined over the set X with orthonormal basis $\{\phi_i\}_{i=1}^n$, and let $K_\phi(\Lambda)$ be the kernel matrix for a reproducing kernel basis $\{k_i(x)\}_{i=1}^m$. Then $\{k_i(x)\}_{i=1}^m$ is a **nearly tight reproducing kernel basis** when*

$$\operatorname{cond}(K_\phi(\Lambda)) \approx 1$$

*and the set $\Lambda = \{\lambda_i\}_{i=1}^m$ then forms a **nearly extremal spanning point set**.*

Nearly tight reproducing kernel bases provide in practice the same benefits than tight ones, for example

$$K_\phi^{-1}(\Lambda) \approx K_\phi^T(\Lambda),$$

and the error acceptable in applications determines when the approximate equality in Def. 4.4 is satisfied. Compared to Proposition 4.5, the notion of a nearly tight reproducing kernel basis is more useful since such representations can usually be constructed by numerically optimizing the reproducing points until a local minimum of the condition number has been reached. We will employ and develop this idea in subsequent sections.

Remark 4.6. The overcompleteness or redundancy of a representation can be interpreted as a measure for the "coverage" of a vector space, which for a basis is critical or just sufficient to span it, cf. Fig. 2.3. One can hence expect that overcompletness improves the condition number $\operatorname{cond}(K_\phi(\Lambda))$, and we will see that this is indeed the case in the following.

4.2 Finitary Point Functionals

In this section, we will employ reproducing kernel bases to develop finitary point functionals. We will also provide examples of such functionals, and analyze the additional insights that can be obtained with our formulation.

4.2.1 Finitary Point Functionals

Using the reproducing kernel bases developed in the last section, we will in the following introduce finitary point functionals as a common perspective for a wide range of computational techniques. Let us begin by formally defining what we mean by a finitary point functional.

Definition 4.5. *Let $\mathcal{H}(X)$ be a separable reproducing kernel Hilbert space defined over the set X. A **finitary point functional** $\mathcal{P} : \mathcal{H}(X) \to \mathcal{G}$ is a computational technique that is formulated using a reproducing kernel basis and which employs only local information of an element in $\mathcal{H}(X)$ at a countable Λ set of locations to determine the image in $\mathcal{G}$.*

Local information in the above definition means that only values from arbitrarily small open neighborhoods are employed, and the use of reproducing kernel bases ensures that this leads to well defined computational techniques.[10] In many applications, in particular those confined to a computer, this is the only type of information available: the only operation that can be performed is evaluating a function. For light transport simulation, for example, the light energy density $\ell \in \mathrm{Den}(T^*Q)$ can only be evaluated locally along individual trajectories in phase space, and one encounters similar situations for example in machine learning and when one has measurements of real world data. The notion of 'finitary' in Def. 4.5 refers to the use of a countable set of locations,[11] and this enables a well defined transition to a computationally viable technique using the classical approach discussed in Chapter 2.2.4.1, which for a reproducing kernel basis amounts to employing a finite subset of reproducing points. This has to be contrasted with more general point functionals such as Monte Carlo integration where no precise interpretation with finite information exists, cf. Fig. 4.2. As we will see shortly, the image space $\mathcal{G}$ in Def. 4.5 can for example be $\mathbb{R}$, as in the case of integration, or again $\mathcal{H}(X)$, as for reconstruction techniques such as sampling theorems and interpolation schemes.

The key to the practicality of finitary point functionals are basis representations whose expansion coefficients are given by local information: reproducing kernel bases as introduced in the last section. As we developed in detail there, the basis functions of these representations are given by the point evaluation functional, when identified with a function in $\mathcal{H}(X)$, and this also motivated our use of 'point functional', see also again Fig. 4.2.

---

[10]In the literature, this is sometimes denoted as 'standard information' (Novak and Woźniakowski, *Tractability of Multivariate Problems: Standard Information for Functionals*).

[11]Our usage of the term 'finitary' was inspired by those by Tao, cf. (*Structure and Randomness: Pages from Year One of a Mathematical Blog*).

infinitary	point functionals	Monte Carlo techniques (probabilistic)

separable reproducing kernel Hilbert space	finitary point functionals	Shannon sampling theorem
structurally equivalent		
finite reproducing kernel Hilbert space	finite point functionals	Gauss-Legendre quadrature Lagrange interpolation
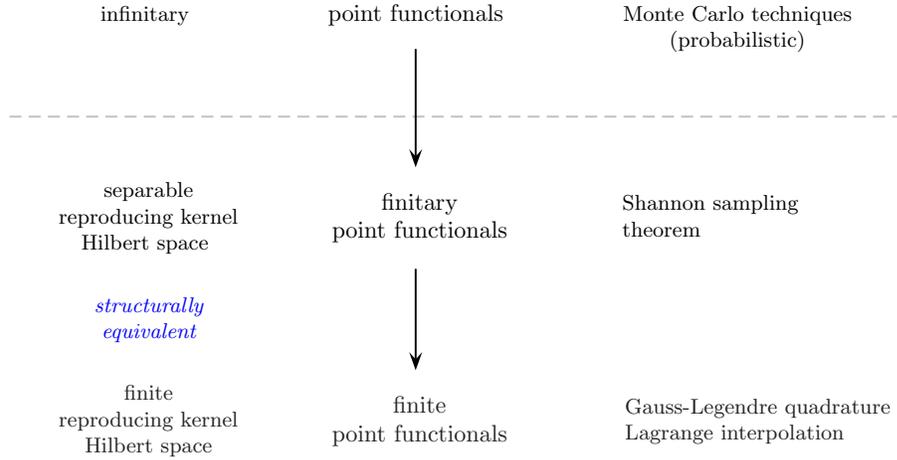
Figure 4.2: Different types of point functionals, that is computational techniques that only employ local information. Note that finitary and finite point functionals are structurally equivalent, since they are both defined in reproducing kernel Hilbert spaces.

Rather than studying the consequences of Def. 4.5 abstractly, we will consider three finitary point functionals that are of importance in applications—and which with our formulation will all be variations of a common theme. A finite setting will thereby often be the natural choice from the outset, making the techniques immediately amenable to computations. We will also continue to develop our ideas without too many digressions, and commentary and the relationship to work in the literature will mostly be relegated to Chapter 4.2.4. $\mathcal{H}(X)$ will again denote a reproducing kernel Hilbert space defined over the set X with orthonormal basis $\{\phi_i\}_{i=1}^n$, and $\{k_i(x)\}_{i=1}^m$ will be a reproducing kernel basis for the reproducing points $\Lambda = \{\lambda_i\}_{i=1}^n$ with dual basis $\{\tilde{k}_i(x)\}_{i=1}^m$.

4.2.1.1 Sampling

Sampling theorems establish when a function can be reconstructed at an arbitrary point from its values at a countable set of locations, with sampling referring to the process of obtaining the values, classically from measurements.[12] The best known example of a sampling theorem is those by Shannon, which establishes a reconstruction formula for functions on the real line that are bandlimited in the Fourier domain, and we will consider this setting in detail in Chapter 4.2.2.4.

---

[12]The term 'sampling' is sometimes used in a more general sense, for arbitrary inner products that yield basis function coefficients, cf. for example (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*), but we will employ it in the more classical sense where point values are employed.

Using reproducing kernel bases and Proposition 4.1, a sampling theorem for a reproducing kernel Hilbert space $\mathcal{H}(X)$ over an arbitrary set X is provided by

$$f(x) = \sum_{i=1}^{m} f(\lambda_i)\, \tilde{k}_i(x) \tag{4.6}$$

and the samples $\mathrm{f}(\Lambda) = (f(\lambda_1), \ldots, f(\lambda_n))$ are the basis function coefficients with respect to the dual kernel basis functions $\tilde{k}_i(x)$, which in the literature are classically denoted as reconstruction filters. An interesting extension of Proposition 4.1 is provided by derivative sampling. Exploiting the linearity of the derivative and following the derivation in Eq. 2.39 backwards, one obtains for the derivative kernel

$$k'_y(x) = \sum_{i=1}^{m} \phi_i(x)\, \phi'_i(y) \tag{4.7}$$

where $\phi'_i(y)$ denotes the derivative $d\phi(y)/dy$. An expansion analogous to Eq. 4.6 is then given by

$$f'(x) = \sum_{i=1}^{m} \langle f'(x), k'_i(x)\rangle\, \tilde{k}'_i(x) = \sum_{i=1}^{m} f'(\lambda_i)\, \tilde{k}'_i(x) \tag{4.8}$$

and the elements of the derivative reproducing matrix $\mathrm{K}'_\phi(\Lambda)$ with respect to the orthonormal basis $\{\phi_i\}_{i=1}^{n}$ are thus

$$k'_{ij} = \phi'_j(\lambda_i). \tag{4.9}$$

It is important to note that by the linearity of the derivative, the values $f'(x)$ also satisfy

$$f'(x) = \sum_{i=1}^{m} f_i(\phi)\, \phi'_i(x) \tag{4.10}$$

and since the same basis function coefficients $f_i(\phi)$ describe function values and derivative values one can employ either of them, or combine them, to reconstruct the $f_i(\phi)$ as in Corollary 4.2. However, with derivative values some care is required since the derivative basis functions $\phi'_i(x)$ are in general no longer normalized. We will return to this question in the next section and in Chapter 4.2.3.2 in the context of a concrete application.

4.2.1.2 Approximation and Interpolation

Scattered data approximation aims at the reconstruction of a continuous function from values at an unstructured set of locations, for example from measurements of real world data. An approximation formula for scattered data is again given by Proposition 4.1, that is

$$f(x) = \sum_{i=1}^{m} f(\lambda_i)\, \tilde{k}_i(x) \tag{4.11}$$

and where it is assumed that the given locations λ_i form a spanning point set. Moreover, when the reproducing kernel basis is a Riesz basis and not overcomplete, then the approximation is interpolatory by Corollary 4.1. For the effectiveness of scattered data approximation in applications usually the choice of a suitable reproducing kernel Hilbert space that corresponds to the signal properties is vital.

An approximation problem of central importance for many applications is the projection of a signal into a basis, mainly because once a basis representation is available many other question are easily formulated, cf. Chapter 2.2.4. From Corollary 4.2, we immediately have for the basis function coefficients $f_i(\phi)$ with respect to an orthonormal basis $\{\phi_i\}_{i=1}^n$ that

$$f_i(\phi) = \sum_{j=1}^m r_{ij}\, f(\lambda_j) \tag{4.12}$$

where the r_{ij} are the elements of the reconstruction matrix $\mathrm{R}_\phi(\Lambda)$ for the reproducing kernel basis $\{k_i(x)\}_{i=1}^m$ defined by the reproducing locations $\Lambda = \{\lambda_i\}_{i=1}^m$.

In most applications, the input signal f is not perfectly contained in the space $\mathcal{H}(X)$, and using Eq. 4.12 will then incur some error, known as aliasing error in the classical literature, which arises from the signal component outside of the space. In fact, in applications $\mathcal{H}(X)$ will often be an approximation space, in the sense of Def. 2.51, and Eq. 4.12 is employed to obtain a finite approximation of an input signal for which only value at a discrete set of locations are known. For these situations, we have the following result.

Proposition 4.6. *Let $\mathcal{G}(X)$ be a Hilbert space with orthonormal basis $\{\phi_i\}_{i=1}^p$, with p possibly being infinity, and $\mathcal{H}(X) \subset \mathcal{G}(X)$ a finite dimensional reproducing kernel Hilbert such that $\{\phi_i\}_{i=1}^n$ forms an orthonormal basis for the space with reproducing kernel basis $\{k_i(x)\}_{i=1}^m$ defined over reproducing points $\Lambda = \{\lambda_i\}$. Moreover, let $f \in \mathcal{G}(X)$ be a function of the form $f = \hat{f} + \check{f}$ where $\hat{f} \in \mathcal{H}(X)$ and $\check{f} \in \mathcal{G}(X)\backslash\mathcal{H}(X)$. The error $err_k(f)$ in the kth basis function coefficient $f_k(\phi) = \langle f, \phi_k \rangle$ when obtained by the basis projection in Eq. 4.12 is then bounded by*

$$|err_k(f)| \leq \|\check{f}\|\|\gamma_k\|$$

where the vector $\gamma_k = (\gamma_{n+1}^k, \ldots, \gamma_p^k)$ has elements

$$\gamma_i^k = \sum_{j=1}^m \phi_i(\lambda_j)\, r_{kj}.$$

By the above proposition, the error is characterized by a signal dependent term $\|\check{f}\|$, which would completely determine the error when Eq. 4.12 would provide the orthogonal projection onto $\mathcal{H}(X)$, and a term that only depends on the basis $\{\phi_i\}_{i=1}^n$ and the reproducing points $\Lambda = \{\lambda_j\}_{j=1}^m$ that are employed. The term $\|\check{f}\|$ is small for example for Sobolev-like spaces where the basis

function coefficients decay sufficiently fast, cf. Example 2.13, and we will see practical examples of such signals in Chapter 4.2.3.2. For the second term $\|\gamma_k\|$, our result is constructive in that it characterizes locations λ_j that reduce or minimize the error, and it can be controlled for instance by choosing the reproducing points λ_j as the zero crossings of the basis functions that span the residual signal component $\check{f}$. This is possible for families of functions with nested zero crossings, such as Fourier-like bases or classical polynomials, and it is in fact employed in Gauss-Legendre quadrature to increase the degree of accuracy, see Chapter 4.2.2.2. For arbitrary bases without such a structure, it is again possible to numerically optimize locations such that the error term $\|\gamma_k\|$ is minimized, and we will see that oversampling has the same effect in Chapter 4.2.3.2.

The error analysis in Proposition 4.6 carries over to derivative sampling, that is when derivative values are employed to determine the basis projection, cf. Chapter 4.2.1.1. As we remarked before, however, some care is required since the derivative basis functions $\phi_i'(x)$ are no longer normalized. For wavelet- or Fourier-like bases, where the oscillations of the basis functions increase with increasing n, the signal component not contained in $\mathcal{H}(X)$ becomes then significantly larger.[13] Again, we will see examples of this situation in Chapter 4.2.3.2.

Remark 4.7. The error characterization in Proposition 4.6 is not restricted to basis projection. It applies to finitary point functionals more generally since these are defined using a reproducing kernel basis, and it can also be employed for linear operators acting on the representations.

Proof of Proposition 4.6.[14] Let $f(\lambda_i)$ be given values of $f \in \mathcal{G}(X)$ at the m locations $\lambda_i \in X$. With Eq. 4.12, we then obtain for the kth basis function coefficient $f_k \equiv f_k(\phi)$ with respect to ϕ_k that

$$f_k = \langle f, \phi_k \rangle \tag{4.13a}$$

$$= \left\langle \sum_{j=1}^{m} f(\lambda_j)\,\tilde{k}_j, \phi_k \right\rangle \tag{4.13b}$$

and since $f = \hat{f} + \check{f}$ we can write

$$f_k = \left\langle \sum_{j=1}^{m} \left(\hat{f}(\lambda_j) + \check{f}(\lambda_j) \right) \tilde{k}_j, \phi_k \right\rangle. \tag{4.13c}$$

---

[13] For the Fourier basis this is also easily seen by the well known formula $\hat{f}'(\zeta) = i\zeta\hat{f}(\xi)$, where $\hat{f}(\xi)$ denotes the Fourier transform of the function $f(x)$.

[14] A similar error analysis can for example be found in (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*), see in particular Chapter 5.1.4., although our result was obtained independent from the literature.

Exploiting linearity then yields

$$f_k = \underbrace{\left\langle \sum_{j=1}^{m} \hat{f}(\lambda_j)\tilde{k}_j \, , \, \phi_k \right\rangle}_{\hat{f}_k(\phi)} + \underbrace{\left\langle \sum_{j=1}^{m} \check{f}(\lambda_j)\tilde{k}_j \, , \, \phi_k \right\rangle}_{\mathrm{err}_k(f)} \tag{4.13d}$$

where $\hat{f}_k(\phi)$ is the exact basis function coefficient, corresponding to the orthogonal projection of the input signal, and the second term $\mathrm{err}_k(f)$ represents the error which is caused by an input not contained in $\mathcal{H}(X)$. Exploiting linearity we obtain for the error term

$$\mathrm{err}_k(f) = \left\langle \sum_{j=1}^{m} \check{f}(\lambda_j) \, \tilde{k}_j \, , \, \phi_k \right\rangle \tag{4.14a}$$

$$= \sum_{j=1}^{m} \check{f}(\lambda_j) \left\langle \tilde{k}_j \, , \, \phi_k \right\rangle \tag{4.14b}$$

$$= \sum_{j=1}^{m} \check{f}(\lambda_j) \, r_{kj} \tag{4.14c}$$

where the r_{kj} are the elements of the reconstruction matrix $\mathrm{R}_\phi(\Lambda)$. With the expansion of the residual signal $\check{f}$ in the basis $\{\phi_i\}_{i=1}^{p}$ for $\mathcal{G}(X)$ one obtains

$$\mathrm{err}_k(f) = \sum_{j=1}^{m} \left(\sum_{i=n+1}^{p} \check{f}_i(\phi) \, \phi_i(\lambda_j) \right) r_{kj} \tag{4.14d}$$

and by interchanging the order of the summations this is equivalent to

$$\mathrm{err}_k(f) = \sum_{i=n+1}^{p} \check{f}_i(\phi) \underbrace{\sum_{j=1}^{m} \phi_i(\lambda_j) \, r_{kj}}_{\gamma_i^k} \,. \tag{4.14e}$$

The error term can hence be written concisely as

$$\mathrm{err}_k(f) = \sum_{i=n+1}^{p} \check{f}_i(\phi) \, \gamma_i^k. \tag{4.14f}$$

With the results from the next section, the elements r_{kj} of the reconstruction matrix can be interpreted as the integration weights for the projection of $\phi_i(\lambda_j)$ with $i > n$ onto the kth basis function $\phi_k(x)$ for $\mathcal{H}(X)$, and γ_i^k thus represents the error of the "projection" using the point samples λ_j as integration nodes, which would vanish were it exact since $\phi_i \notin \mathcal{H}(X)$ for $i > n$.

The effect of the signal and the coefficient γ_i^k on the error $\mathrm{err}_k(f)$ can be separated using the Cauchy-Schwarz inequality from Theorem 2.4. One then obtains

$$|\mathrm{err}_k(f)| \leq \|\check{f}\| \|\gamma_k\| \tag{4.15}$$

where $\check{f}$ is the vector of basis function coefficients beyond the bandlimit and γ_k the corresponding vector of the γ_i^k. Eq. 4.15 is again the error characterization provided in Proposition 4.6, and it hence concludes our proof. $\qquad\square$

4.2.1.3 Integration

Integration is arguably the most important linear functional and it is central to many applications in computational science and engineering. An integration rule which is exact for all functions $f \in \mathcal{H}(X)$ in a reproducing kernel Hilbert space $\mathcal{H}(X)$ is obtained from Proposition 4.1 using the linearity of the integral,

$$\int f(x)\,dx = \int \sum_{i=1}^{m} f(\lambda_i)\,\tilde{k}_i(x)\,dx \tag{4.16a}$$

$$= \sum_{i=1}^{m} f(\lambda_i) \int \tilde{k}_i(x)\,dx. \tag{4.16b}$$

This is conveniently written as

$$\int f(x)\,dx = \sum_{i=1}^{m} w_i\,f(\lambda_i) \tag{4.17a}$$

where the weights w_i are defined by

$$w_i = \int \tilde{k}_i(x)\,dx = \sum_j r_{ji} \int \phi_j(x)\,dx. \tag{4.17b}$$

For a closed subspace of $L_2(X)$, where the inner product is given by an integral, the basis projection of Chapter 4.2.1.2 can also be considered as an integration rule with the quadrature weights for the jth basis function coefficient being the elements r_{ij} of the reconstruction matrix $\mathrm{R}_\phi(\Lambda)$. By definition we have

$$\langle f(x), \phi_j(x) \rangle = \int_X f(x)\,\phi_j(x)\,dx \tag{4.18a}$$

and expanding $f(x)$ in the reproducing kernel basis yields

$$\langle f(x), \phi_j(x) \rangle = \int_X \left(\sum_{i=1}^{m} f(\lambda_i)\,\tilde{k}_i(x) \right) \phi_j(x)\,dx. \tag{4.18b}$$

Interchanging summation and integration yields

$$\langle f(x), \phi_j(x) \rangle = \sum_{i=1}^{m} f(\lambda_i) \int_X \tilde{k}_i(x)\,\phi_j(x)\,dx \tag{4.18c}$$

and by definition of the reconstruction matrix we have

$$\langle f(x), \phi_j(x) \rangle = \sum_{i=1}^{m} f(\lambda_i)\,r_{ij} \tag{4.18d}$$

which is obviously just Corollary 4.2.

An integral can be formulated using the basis expansion of the integrand, and hence the error analysis of Proposition 4.6 can be employed when a residual signal component outside of $\mathcal{H}(X)$ exists. The error of the integration scheme in Eq. 4.17 for a function contained in $\mathcal{H}(X)$ can be characterized by a generalized Koksma-Hlawka inequality.

Proposition 4.7. *Let $\mathcal{H}(X)$ be a finite dimensional reproducing kernel Hilbert space of dimension n defined over the set X, and let $\{k_i(x)\}_{i=1}^m$ be a reproducing kernel basis defined over the reproducing points $\Lambda = \{\lambda_i\}_{i=1}^m$. Then the error of the integration rule in Eq. 4.17 for any $f \in \mathcal{H}(X)$ is bounded by*

$$E_m(f) \leq \|f\| \, \mathcal{D}(\Lambda)$$

where the discrepancy $\mathcal{D}(\Lambda)$ of a set of locations Λ is

$$\mathcal{D}(\Lambda) = C_p \, C_{\mathrm{K}} \, \mathrm{cond}(\mathrm{K}_\phi(\Lambda))$$

and the constants C_p and C_K are independent of Λ.

The above proposition again justifies our use of the condition number as a quality measure for reproducing kernel bases.

Remark 4.8. The classical Koksma-Hlawka inequality bounds the error $E_m^{\mathrm{BV}}(f)$ of an m-point Quasi Monte Carlo integration, or equal weight quadrature, in the space $\mathrm{BV}(X)$ of functions of bounded variation $V(f) < \infty$, cf. Example 2.14. The inequality is given by

$$E_m^{\mathrm{BV}}(f) \leq V(f) \, \mathcal{D}(\Lambda)$$

where $\mathcal{D}(\Lambda)$ is the discrepancy of a sequence $\Lambda = \{\lambda_i\}$ of points λ_i, typically a low discrepancy sequence.[15]

Proof of Proposition 4.7.[16] The error of the integration rule in Eq. 4.17a is naturally defined as

$$E_m(f) = \left\| \int_X f(x) \, dx - \sum_{i=1}^m w_i \, f(\lambda_i) \right\|, \tag{4.19}$$

---

[15]See the classic text by Niederreiter (*Random Number Generation and Quasi-Monte Carlo Methods*) for more details on low discrepancy sequences and the various notions of discrepancy. Interestingly, it was again Weyl ("Über die Gleichverteilung von Zahlen mod. Eins"), whom we encountered before in the context of geometric mechanics, who introduced the notion of low discrepancy sequences.

[16]Our derivation follows closely the recent monograph by Dick and Pillichshammer (*Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Chapter 2) for more general reproducing kernel Hilbert spaces, and the idea to employ reproducing kernels to characterize discrepancies goes back to (Hickernell, "Quadrature Error Bounds with Applications to Lattice Rules").

With foresight, the integral in the above equation can be written as

$$\int_X f(x)\,dx = \int_X \langle f(y), k_x(y)\rangle\,dx \tag{4.20a}$$

$$= \left\langle f(y), \int_X k_y(x)\,dx \right\rangle \tag{4.20b}$$

and for the integration rule one can write

$$\sum_{i=1}^m w_i\,f(\lambda_i) = \sum_{i=1}^m w_i\,\langle f(y), k_i(y)\rangle \tag{4.21a}$$

$$= \left\langle f(y), \sum_{i=1}^m w_i\,k_i(y) \right\rangle. \tag{4.21b}$$

Using Eq. 4.20 and Eq. 4.21 we obtain for Eq. 4.19 that

$$E_m(f) = \left\| \left\langle f(y), \int_X k_y(x)\,dx \right\rangle - \left\langle f(y), \sum_{i=1}^m w_i\,k_i(y) \right\rangle \right\| \tag{4.22a}$$

and by the linearity of the inner product this is equivalent to

$$E_m(f) = \left\| \left\langle f(y), \int_X k_y(x)\,dx - \sum_{i=1}^m w_i\,k_i(y) \right\rangle \right\|. \tag{4.22b}$$

The Cauchy-Schwarz inequality from Theorem 2.4 then yields

$$E_m(f) \le \|f(y)\| \left\| \int_X k_y(x)\,dx - \sum_{i=1}^m w_i\,k_i(y) \right\| \tag{4.23}$$

which is our generalized Koksma-Hlawka inequality for integration in finite reproducing kernel Hilbert spaces. The generalized discrepancy $\mathcal{D}_m^{\mathcal{H}}(\Lambda)$, which, as one would expect, only depends on the function space $\mathcal{H}(X)$ under consideration and the locations Λ, is therefore

$$\mathcal{D}_m^{\mathcal{H}}(\Lambda) = \left\| \int_X k_y(x)\,dx - \sum_{i=1}^m w_i\,k_i(y) \right\|. \tag{4.24}$$

With the definition of the integration weights in Eq. 4.17b, we can bound the above discrepancy by a more practical expression. For fixed y, Eq. 4.24 is the integration error for the reproducing kernel function $k_y(x)$.[17] Expanding the definition of the weights yields

$$w_i = \int_X \tilde{k}_i(x)\,dx = \sum_{j=1}^n r_{ji} \int_X \phi_j(x)\,dx = \sum_{j=1}^n p_j\,r_{ji} \tag{4.25}$$

---

[17] This quadrature error is intimately related to the worst-case integration error in a reproducing kernel Hilbert space, cf. (Novak and Woźniakowski, *Tractability of Multivariate Problems: Standard Information for Functionals*).

where the r_{ji} are the elements of the reconstruction matrix $\mathrm{R}_\phi(\Lambda)$ and p_j is the integral of the jth basis function ϕ_j. Substituting Eq. 4.25 in Eq. 4.24 we obtain

$$\mathcal{D}_m^{\mathcal{H}}(\Lambda) = \left\| \int_X k_y(x)\,dx - \sum_{i=1}^m \left(\sum_{j=1}^n p_j\, r_{ji} \right) k_i(y) \right\| \qquad (4.26\mathrm{a})$$

and changing the order of the summations yields

$$\mathcal{D}_m^{\mathcal{H}}(\Lambda) = \left\| \int_X k_y(x)\,dx - \sum_{j=1}^n p_j \sum_{i=1}^m r_{ji}\, k_i(y) \right\|. \qquad (4.26\mathrm{b})$$

In matrix-vector notation the last equation becomes

$$\mathcal{D}_m^{\mathcal{H}}(\Lambda) = \left\| \int_X k_y(x)\,dx - \sum_{j=1}^n p_j\, (\mathrm{R}(\Lambda)\,\mathrm{k}_y(\Lambda))_j \right\| \qquad (4.26\mathrm{c})$$

where $\mathrm{k}_y(\Lambda) = (k_y(\lambda_i), \ldots, k_y(\lambda_m))$ and $(\mathrm{R}(\Lambda)\,\mathrm{k}_y(\Lambda))_j$ is the jth element of the vector defined by the matrix vector product $\mathrm{R}(\Lambda)\,\mathrm{k}_y(\Lambda)$. Using the triangle and the Cauchy-Schwarz inequality to separate the terms, the integration error for $k_y(x)$ can, up to a constant C_p which only depends on the integrals p_j of the basis functions ϕ_j,[18] be defined through the error in the solution of the linear system

$$\mathrm{R}\,\mathrm{k}(\Lambda) = \mathrm{K}^{-1}(\Lambda)\,\mathrm{k}(\Lambda). \qquad (4.27)$$

With standard results from linear algebra,[19] the discrepancy is hence bounded by

$$\mathcal{D}(\Lambda) \equiv \mathcal{D}_m^{\mathcal{H}}(\Lambda) \leq C_p\, C_{\mathrm{K}}\, \mathrm{cond}(\mathrm{K}) \qquad (4.28)$$

where the constant C_K is the relative error in $\mathrm{K}(\Lambda)$ and $\mathrm{k}(\Lambda)$, and which for our purposes can be bounded independent of the locations Λ.

$\square$

4.2.2 Classical Point Functionals

In the following, by making choices for the reproducing kernel Hilbert space and the reproducing points that were left abstract in the foregoing derivations, we will recover classical techniques from the literature using our formulation of finitary point functionals.

---

[18]For bases where the integral of all but one basis functions vanishes, which include for example Fourier bases, Legendre polynomials, and many wavelets, the Chebychev sum inequality can be used to obtain an explicit expression for the constant C_p.

[19]Golub and Van Loan, *Matrix Computations*, p. 25.

4.2.2.1   Monte Carlo Integration

Monte Carlo integration is a widely integration rule when little is known about the properties of the integrand. It is usually justified using probabilistic arguments,[20] but we will provide in the following a functional analytic interpretation.

Let $\Xi_n(X) = \{\chi_i\}_{i=1}^n$ be a characteristic basis defined over a uniform partition $P = \{P_i\}_{i=1}^n$ for $X = [a, b] \subset \mathbb{R}$, see Chapter 2.2.5.1. With one reproducing location λ_i in each P_i, the χ_i form an orthogonal reproducing kernel basis, cf. Eq. 2.55, and since the partition is uniform so that $|P_i| = |P_j|$, the weights for the integration rule in Eq. 4.17 are given by

$$w_i = |P_i| = |X|/n \tag{4.29}$$

and independent of χ_i. The integration rule for the space Ξ_n spanned by the characteristic functions is thus

$$\int f(x)\,dx = \sum_{i=1}^n \frac{|X|}{n} f(\lambda_i) = \frac{|X|}{n} \sum_{i=1}^n f(\lambda_i) = \frac{b-a}{n} \sum_{i=1}^n f(\lambda_i) \tag{4.30}$$

which is the standard Monte Carlo estimator for uniformly distributed sampling locations. For samples drawn from an arbitrary probability distribution function $p(x) : \mathbb{R} \to \mathbb{R}$, the requirement of one sample in the support of every unit height characteristic basis functions χ_i implies that the partition can no longer be uniform but that the P_i have to have the form $P_i = [x_i, x_{i+1}]$ for suitable bounds $x_i \in [a, b]$. Choosing the bounds x_i such that in the support of every χ_i is on average one sample is equivalent to

$$P([x_i, x_{i+1}]) = \int_{x_i}^{x_{i+1}} p(x)\,dx = 1/n. \tag{4.31a}$$

With arbitrary $y \in [x_i, x_{i+1}]$, this is to zero[th] order equivalent to

$$n\,(p(y)\,(x_{i+1} - x_i)) = 1 \tag{4.31b}$$

and with y being the sample λ_i in the support of χ_i one thus has for the size of the partition P_i that

$$|P_i| = x_{i+1} - x_i = \frac{1}{n\,p(\lambda_i)}. \tag{4.32}$$

The integration rule for samples distributed according to $p(x)$ is therefore

$$\int f(x)\,dx = \sum_{i=1}^n \frac{1}{n\,p(\lambda_i)} f(\lambda_i) = \frac{1}{n} \sum_{i=1}^n \frac{f(\lambda_i)}{p(\lambda_i)} \tag{4.33}$$

which coincides with the standard Monte Carlo estimator when importance sampling according to $p(x)$ is employed.

---

[20]See for example (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*, Chapter 13) for a classical introduction to Monte Carlo integration.

By Theorem 2.10, $\Xi_n(X)$ becomes dense in $L_2(X)$ as n goes to infinity, which recovers the well known result that Monte Carlo integration converges for any function in $L_2(X)$ ad infinitum. Note that in the limit also the zero[th] order approximation in Eq. 4.31 becomes accurate. The above derivation is also easily extended from the real line to more general measure spaces (X, Σ, μ), since it only relies on the existence of a suitable partition $P = \{P_i\}_{i=1}^n \subset \Sigma$.

4.2.2.2 Gauss-Legendre Quadrature[21]

Gauss-Legendre quadrature rules are defined for the spaces $\mathcal{P}_{n-1}([-1,1])$ spanned by all Legendre polynomials $P_l(x)$ up to degree $n-1$. The integrals of all Legendre polynomials except P_0 vanish. The weights for the integration rule in Eq. 4.17 are thus $w_i = r_{0i}$ and given by the first row of the reconstruction matrix $\mathrm{R}_{P_l}(\Lambda)$. From the point of view of approximation theory, this results follows from the fact $P_0(x)$ is the constant function, so that the first basis function coefficient is, up to a constant, the sought integral. Choosing the zero crossings of P_n as nodes for the integration rule for $\mathcal{P}_{n-1}([-1,1])$, which as we discussed in Chapter 4.2.1.3 avoids error by functions in $\mathcal{P}_n([-1,1])$, then recovers classical Gauss-Legendre quadrature.[22]

4.2.2.3 Lagrange Interpolation[23]

A classical method for the interpolation of scattered data is Lagrange interpolation which is based on the monomial basis $\{1, x, \ldots, x^n\}$ over $X = [-1, 1]$. With a set of reproducing points $\Lambda = \{\alpha_i\}$ at m locations $\alpha_i \in [-1, 1]$, the reproducing matrix is

$$\mathrm{V}(\Lambda) \equiv \mathrm{K}(\Lambda) = \begin{bmatrix} 1 & \ldots & \alpha_1^n \\ \vdots & \ddots & \vdots \\ 1 & \ldots & \alpha_m^n \end{bmatrix} \in \mathbb{R}^{m \times n}. \tag{4.34}$$

In the literature, $\mathrm{V}(\Lambda)$ is known as Vandermonde matrix, and it is nonsingular if and only if the locations α_i are distinct. With our ansatz, the dual kernel functions $\ell_i(x) = \tilde{k}_i(x)$ are obtained by computing the inverse of $\mathrm{V}(\Lambda)$ and expanding the dual basis function coefficients in the monomial reference basis. A closed form expression for the expansion is given by

$$\ell_j(x) = \prod_{i=1, i \neq j}^{m} \frac{x - \alpha_i}{\alpha_j - \alpha_i} \tag{4.35}$$

---

[21]See for example (Press et al., *Numerical Recipes in C: The Art of Scientific Computing*, p. 150) for a classical derivation.

[22]Examples of this construction are available at http://www.dgp.toronto.edu/people/lessig/dissertation/files/gauss-legendre.nb.

[23]See for example (Press et al., *Numerical Recipes in C: The Art of Scientific Computing*, Chapter 3.1).

which is the classical Lagrange interpolation polynomial.[24] The discrete Fourier transform matrix is sometimes also denoted as Vandermonde matrix, and in our parlance it is the reconstruction matrix for a bandlimited Fourier series that recovers the Fourier coefficients from samples at regularly spaced locations in $[-\pi, \pi]$, although our formulation also easily accommodates arbitrarily spaced samples.

4.2.2.4 Shannon Sampling Theorem[25]

The setting of the Shannon sampling theorem is the Paley-Wiener space $\Omega_B(\mathbb{R})$ of Fourier-bandlimited functions over the real line $\mathbb{R}$ for the orthonormal Fourier functions $\phi(\xi, x)$ with frequency $\xi \in \mathbb{R}$. For a bandlimited function $f(x) \in \Omega_B$ with frequency representation $\hat{f}(\xi)$, one thus has[26]

$$f(x) = \int_{-B}^{B} \hat{f}(\xi) \, \phi(\xi, x) \, d\xi \tag{4.36a}$$

$$= \int_{-B}^{B} \left(\int_X f(y) \, \phi(\xi, y) \, dy \right) \phi(\xi, x) \, d\xi \tag{4.36b}$$

and by Fubini's theorem this can be written as

$$f(x) = \int_X f(y) \left(\int_{-B}^{B} \phi(\xi, x) \, \phi(\xi, y) \, d\xi \right) dy. \tag{4.36c}$$

But with

$$\mathrm{sinc}_B(x - y) = \int_{-B}^{B} \phi(\xi, x) \, \phi(\xi, y) \, d\xi, \tag{4.36d}$$

which is just the continuous analogue of the expansion of the reproducing kernel in an orthonormal basis in Proposition 2.36, one obtains

$$f(x) = \int_X f(y) \, \mathrm{sinc}_B(x - y) \, dy. \tag{4.36e}$$

---

[24]For a numerical verification of the above claim see http://www.dgp.toronto.edu/people/lessig/dissertation/files/Lagrange-interpolation.nb.

[25]The reader is reminded of Arnold's principle, which states that a result is never named after its first discoverer, and before Shannon the sampling theorem usually named after him was obtained by Whittaker, Nyquist, Ogura, Kotelnikov, ..., although we believe it is fair to say that it was Shannon who realized the importance of the result. We refer to the literature for a discussion of the history (Butzer et al., "Interpolation and Sampling: E.T. Whittaker, K. Ogura and Their Followers"). Modern treatments of the theorem can for example be found in (Nashed and Walter, "General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces"; Higgins, "Sampling Theory for Paley-Wiener Spaces in the Riesz Basis Setting").

[26]For Eq. 4.36d, see for example (Nashed and Walter, "General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces", Eq. 1.3) and a numerical validation can be found at http://www.dgp.toronto.edu/people/lessig/dissertation/files/sinc-integral-expansion.nb.

This is the sampling step of Shannon's theorem. The theorem, which described the reconstruction of a function from its values $f(k)$ at the integers $k \in \mathbb{Z}$ by

$$f(x) = \sum_{k=-\infty}^{\infty} \langle f(y), \mathrm{sinc}_B(y-k) \rangle \, \mathrm{sinc}_B(x-k) \qquad (4.36\mathrm{f})$$

$$= \sum_{k=-\infty}^{\infty} f(k) \, \mathrm{sinc}_B(x-k), \qquad (4.36\mathrm{g})$$

is obtained by forming a kernel basis using sinc_B-functions. With the integers as sampling locations and exploiting that the zero crossings of the sinc_B-function are $\mathrm{sinc}_B(k-l) = \delta_{kl}$, for $k, l \in \mathbb{Z}$, it is easy to see that the infinite dimensional Gramian of the system is the identity, and the basis hence orthonormal. Note that Fubini's theorem, which was essential in deriving the result, can only be employed when $B < \infty$, which is the classic sampling condition in Shannon's theorem.

4.2.3 Applications

In this section, we will demonstrate the practical relevance of reproducing kernel bases and finitary point functionals, and we will develop two applications where reproducing kernel bases are employed for numerical computations.

4.2.3.1 Rotation of Functions on the Sphere[27]

When reproducing kernel bases are employed, one can work with the values of a function at a set of locations instead of traditional basis function coefficients. In the following, we will employ this insight to obtain an accurate and efficient algorithm for rotating finite spherical harmonics expansions.

Rotation and Sampling Let $f \in \mathcal{H}_L = \mathcal{H}_{\leq L}$ be an L-bandlimited signal in the spherical harmonics domain with basis expansion

$$f(\omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \beta_{lm} \, y_{lm}(\omega) \qquad (4.37)$$

and denote by f_l the component of the signal in the lth band $\mathcal{H}_l$ so that $f = f_0 + \ldots + f_L$. We are interested in the basis function coefficients $\bar{\beta}_{lm}$ of the rotated signal

$$\bar{f} = \mathrm{R}f \qquad (4.38)$$

---

[27]An introduction to spherical harmonics can be found in Chapter 2.2.5.2, and ω will in the following denote a point on the sphere, and not the symplectic 2-form or the angular frequency. The material in this section appeared before in (Lessig, de Witt, and Fiume, "Efficient and Accurate Rotation of Finite Spherical Harmonics Expansions").

for arbitrary $R \in SO(3)$. Individual spherical harmonics bands $\mathcal{H}_l$ are closed under the action of $SO(3)$, and it is hence sufficient to consider an arbitrary but fixed space $\mathcal{H}_l$. As shown in Chapter 2.2.5.2, the reproducing kernel for $\mathcal{H}_l$ is given by

$$k_l(\bar{\omega}, \omega) = \sum_{m=-l}^{l} y_{lm}(\bar{\omega}) \, y_{lm}(\omega) = \frac{2l+1}{4\pi} P_l(\bar{\omega} \cdot \omega) \qquad (4.39)$$

where $P_l(x)$ is the Legendre polynomial of degree l. A reproducing kernel basis $\{k_i(\omega)\}_{i=1}^{m}$ for $\mathcal{H}_l$ can thus be formed using at least $\dim(\mathcal{H}_l) = (2l+1)$ locations $\Lambda \equiv \Lambda_l = \{\lambda_i\}$ on the sphere, and for the moment we will posit that these yields linearly independent functions. With the group $SO(3)$ acting on functions by the pointwise pullback

$$(Rf)(\omega) = f(R^{-1}\omega), \qquad (4.40)$$

we obtain for the representation of the rotated signal $\bar{f}_l = Rf_l$ in the reproducing kernel basis that

$$Rf_l = \sum_{i=1}^{n} \langle Rf_l(\omega), k_i(\omega) \rangle \tilde{k}_i(\omega) \qquad (4.41a)$$

$$= \sum_{i=1}^{n} f_l(R^{-1}\lambda_i) \, \tilde{k}_i(\omega). \qquad (4.41b)$$

The above equation is a sampling theorem for functions in $\mathcal{H}_l$ but with rotated sampling locations. Hence, the basis function coefficients of the rotated signal with respect to the reproducing kernel basis are given by the values of the unrotated signal f_l at rotated sampling locations $\bar{\lambda}_i = R^{-1}\lambda_i$. Exploiting that the reconstruction matrix $R_l(\Lambda)$ provides the change of basis from the kernel basis to spherical harmonics, the sought after coefficients $\bar{\beta}_{lm}$ of the rotated signal are recovered by

$$\bar{\beta}_l = R_l(\Lambda) \, \bar{f}_l(\Lambda) \qquad (4.42)$$

where $\bar{\beta}_l = (\bar{\beta}_{l,-l}, \ldots, \bar{\beta}_{l,l})$, and $\bar{f}_l = (f_l(\bar{\lambda}_1), \ldots, f_l(\bar{\lambda}_n))$ is the signal at the rotated sampling points. Since the $f_l(\bar{\lambda}_i)$ can be evaluated using the basis function coefficients β_{lm} of the unrotated signal, Eq. 4.42 provides a practical algorithm for rotating finite spherical harmonics expansions.

Reproducing Points for the Sphere   For the spherical harmonics bands $\mathcal{H}_l$, the existence of a spanning point set is guaranteed by a theorem due to Müller,[28] and it can be shown that up to a set of measure zero any set of $(2l+1)$ points $\lambda_i \in S^2$ can be employed.[29] However, as we discussed before in

---

[28](Müller, *Spherical Harmonics*, p. 13), (Freeden, Gervens, and Schreiner, *Constructive Approximation on the Sphere (With Applications to Geomathematics)*, p. 51)

[29]This is analogous to the situation for the Vandermonde matrix in Chapter 4.2.2.3.
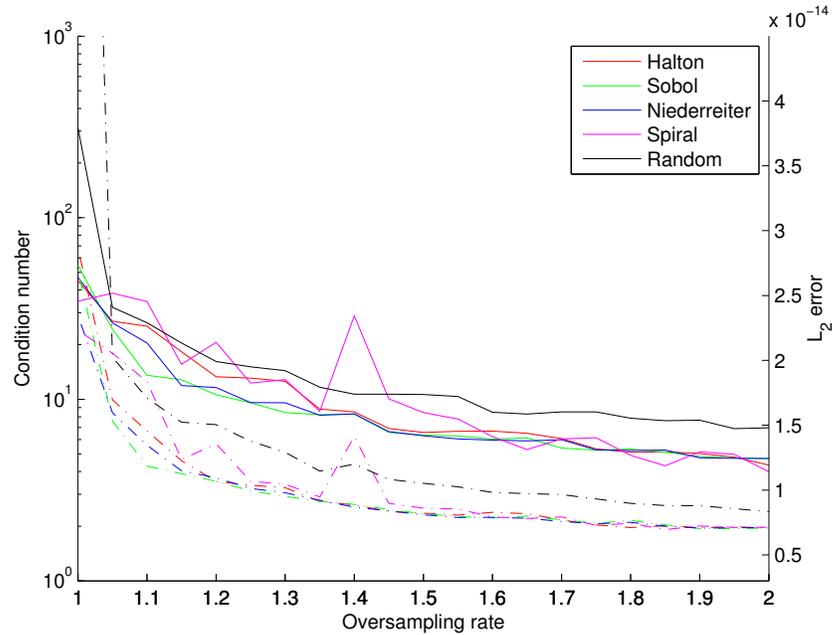
Figure 4.3: Condition number of the reproducing matrix (full) and average L_2 rotation error (dashed) as a function of the oversampling rate for $l = 20$.

Chapter 4.1.3, not only the existence of spanning point sets is of importance but also the quality as described by the condition number of the reproducing matrix, cf. Def. 4.4.

As one would expect, $(2l + 1)$ well distributed points on the sphere yield low condition numbers and are well suited for our rotation algorithm. We obtain such locations by mapping quasi-random sequences from the unit square to the upper hemisphere,[30] the restriction to the hemisphere is advantageous since anitpodal points would yield co-linear basis functions. Additionally, we also employ the spiral points that were proposed by Saff and Kuijlaars[31]. Fig. 4.3 shows that the performance of different well distributed sequences is qualitatively equivalent, and that they outperform sampling points obtained with a (pseudo) random number generator. The graphs also verify the close correlation between condition number and rotation error, a connection that is only violated when a rotated sampling location is close to a pole and the accurate evaluation of spherical harmonics is difficult.

As we suggested previously, the condition number of the reproducing matrix can be improved using overcomplete representations and by numerically optimizing the sampling locations. Fig. 4.3 demonstrates empirically that over-

---

[30]Cui and Freeden, "Equidistribution on the Sphere".

[31]Saff and Kuijlaars, "Distributing many points on a sphere".
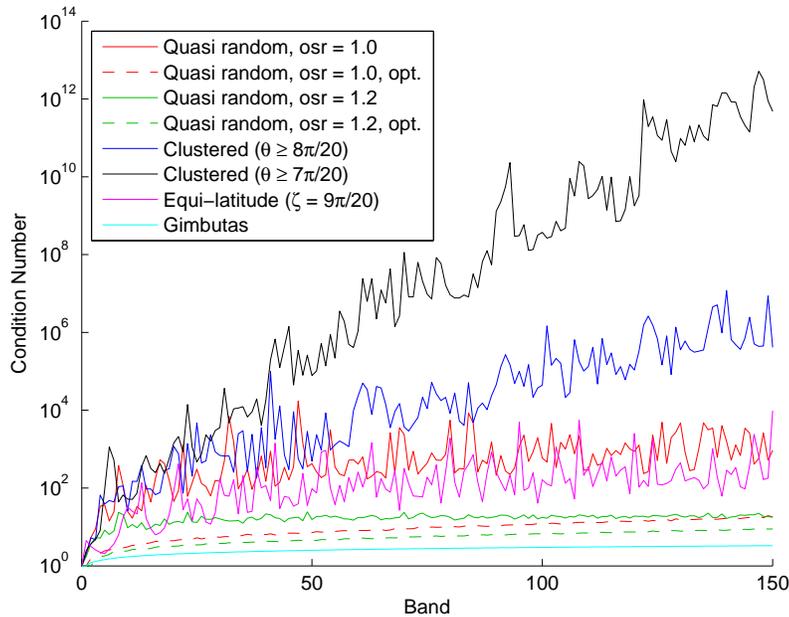
Figure 4.4: Condition numbers of the reproducing matrix for different strategies to obtain reproducing points and with different oversampling rates (osr).

complete representations improve the condition number and the rotation error, and that a small oversampling rate is sufficient to obtain nearly tight kernel bases in the sense of Def. 4.4. Additionally, one can also show that the condition number approaches unity as the number of sampling points goes to infinity. Oversampling requires more function evaluations than critical sampling, making it computationally more expensive. Numerically optimizing the reproducing points allows to avoid this drawback while still retaining the advantages of close to optimal reproducing points. Using a BFGS optimizer with finite differences and restarts based on different well distributed point sets,[32] we obtained considerable improvements in the condition number. For lower bands, combining oversampling and optimization then yielded again nearly tight reproducing kernel bases, see Fig. 4.4. The main limitation of optimization are the high computational costs, although this is necessary only once and can easily be performed in a precomputation step.

Overcomplete representations and the numerical optimization of the reproducing points can be employed for arbitrary function spaces and domains. An interesting choice specific to the sphere is provided by reproducing points $\lambda_i = (\zeta, \phi_i)$ with a fixed latitude ζ. By Eq. 2.56, the reproducing matrix can then be factored as $K_l(\zeta) = P_l(\zeta)\, F$, where F is the discrete Fourier transform matrix and the nonzero elements of the diagonal matrix $P_l(\zeta)$ are given by

[32]Zhu et al., "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization".

$p_{mm} = P_{lm}(\cos \zeta)$ with $-l \leq m \leq l$. This enables to recover the basis function coefficients $\bar{\beta}_{lm}$ of the rotated function by performing a discrete Fourier transform and then dividing by $P_{lm}(\cos \zeta)$. From the orthogonality of the discrete Fourier transform, it follows that the condition number $\mathrm{cond}(\mathrm{K}_l(\zeta))$ of the reproducing matrix for equi-latitude points is given by

$$\mathrm{cond}(P_l(\zeta)) = \frac{\max\left(|P_{lm}(\cos \zeta)|\right)}{\min\left(|P_{lm}(\cos \zeta)|\right)} \tag{4.43}$$

and the accuracy of the algorithm is hence determined by the latitude ζ, cf. Remark 4.4. Experiments show that beyond a critical latitude the condition number deteriorates as the distance from the equator increases, and for $l \leq 150$ latitudes with $75° \leq \zeta < 90°$ should be employed. When other latitudes are used or for very large l, the accuracy is insufficient only for a very small number of rotated coefficients $\bar{\beta}_{lm}$ where $P_{lm}(\zeta)$ is very small. A practical work-around is to compute these coefficients with the algorithm in Eq. 4.42 using a small number of additional sampling points with $\theta \neq \zeta$. An interesting choice for the latitude is $\zeta = \pi/2$, which locates the kernel locations on the equator. The sampling points are then not a spanning point set, but together with derivative sampling this provides a highly accurate algorithm, as was recently shown by Gimbutas and Greengard.[33]

Experimental Evaluation   We evaluated our rotation algorithm with different choices for the sampling locations, and compared its accuracy and performance to various techniques in the literature.[34] As reproducing points we employed optimized point sets with different oversampling rates, nested sampling points where the locations for band l are a subset of those for band $l + 1$, and equi-latitude points with $\zeta = 9\pi/20$ for which the fast Fourier transform was used to speed up computations. With nested sampling points, the recurrent structure of spherical harmonics can be exploited to enable a faster evaluation at the sampling locations. Even with nesting, a large fraction of the computation time is spent on determining spherical harmonics values. We therefore compute the sample values $f_l(\bar{\lambda}_i)$ also from a representation of f_l in the reproducing kernel basis, with primary and dual basis functions interchanged, in which case it suffices to evaluate Legendre polynomials.

From the literature we employed the algorithms by Ivanic and Ruedenberg,[35] Blanco, Florez, and Bermejo,[36] Pinchon and Hoggan,[37] and Gimbutas and

---

[33] Gimbutas and Greengard, "A fast and stable method for rotating spherical harmonic expansions".

[34] A Matlab implementation of our algorithm as well as the C++ framework used for the experiments, including optimized sampling sequences, is available at http://www.dgp.toronto.edu/people/lessig/shrk/. Parts of the Matlab code are based on Frederik Simons repository: http://geoweb.princeton.edu/people/simons/software.html.

[35] Ivanic and Ruedenberg, "Rotation Matrices for Real Spherical Harmonics. Direct Determination by Recursion".

[36] Blanco, Flórez, and Bermejo, "Evaluation of the rotation matrices in the basis of real spherical harmonics".

[37] Pinchon and Hoggan, "Rotation Matrices for Real Spherical Harmonics: General Rotations
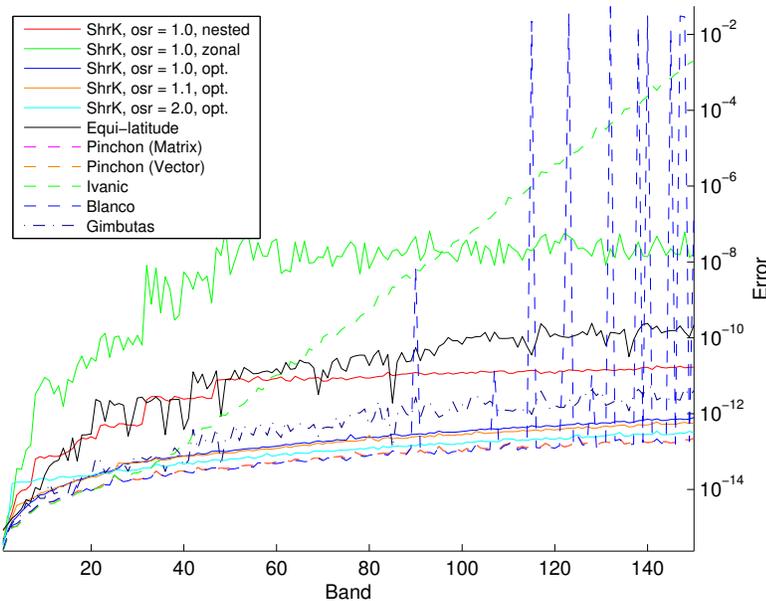
Figure 4.5: Average L_2 error for the rotation of signals on the sphere for various techniques (osr = oversampling rate).

Greengard,[38] implemented in C/C++ using double precision and, when available, based on programs provided by the original authors. For the algorithm proposed by Pinchon and Hoggan we employed two variants:[39] the first obtains the spherical harmonics rotation matrix and then applies it to a coefficient vector as proposed in the original publication ('Pinchon Matrix'), the second is a variation suggested to us by the authors that avoids the explicit computation of the rotation matrix and applies the coefficient vector incrementally to the highly sparse factorization employed in the work ('Pinchon Vector'). The latter version has a computational complexity of $O(L^2)$, whereas all other algorithms in the literature, including ours, have a complexity of at least $O(L^3)$.

Average L_2 errors per spherical harmonics band $\mathcal{H}_l$ are reported in Fig. 4.5. Other error norms are qualitatively equivalent to the presented results and have been omitted. All graphs have been obtained by averaging over a large number of random rotations, and reference solutions were obtained as in previous work.[40] As shown in Fig. 4.5, optimized sampling locations (Shrk, osr = *, opt.) provide significantly higher accuracy than well distributed sequences (ShrK, osr = 1.0, nested) and equi-latitude points, in particular when combined

of Atomic Orbitals in Space-Fixed Axes".

[38]Gimbutas and Greengard, "A fast and stable method for rotating spherical harmonic expansions".

[39]Pinchon and Hoggan, "Rotation Matrices for Real Spherical Harmonics: General Rotations of Atomic Orbitals in Space-Fixed Axes".

[40]Gimbutas and Greengard, "A fast and stable method for rotating spherical harmonic expansions".

with oversampling. In this case only little accuracy is lost and our algorithm performs comparably to the technique by Pinchon and Hoggan.[41]

Discussion   With our algorithm, the spherical harmonics coefficients of a rotated signal are obtained using a sampling formula for the sphere and by exploiting the pointwise action of the rotation group SO(3) on functions. In contrast to techniques in the literature which construct spherical harmonics rotation matrices, our algorithm is simple to implement and enables one to trade-off accuracy and performance, making it well suited for a wide range of applications. We presented extensive numerical experiments to validate our approach and to compare it to the literature, and to our knowledge this is the first thorough evaluation of spherical harmonics rotation algorithms. The experimental results demonstrate that our technique attains accuracy comparable to the best methods. A possible disadvantage of our approach is that the reconstruction matrices R_l have to be precomputed and stored. However, other algorithms suffers from similar drawbacks, and for our technique it can be avoided with equi-latitude sampling locations.

Our algorithm was inspired by work by Higgins and Kempski[42] and Freeden and co-workers[43] who proposed sampling theorems for the sphere similar to Eq. 4.41. However, these authors did not consider biorthogonal and overcomplete representations, which are vital for our technique. Algorithms similar to ours were proposed before by Stern[44] and by Gimbutas and Greengard.[45] However, Stern did not investigate different sampling locations or the mathematics underlying the technique, and his work unfortunately received only very little attention after its inception. Gimbutas and Greengard recently proposed an interesting variation of our technique with sampling points on the equator. Our work provides a general framework for their algorithm and clarifies its theoretical foundations.

4.2.3.2   Basis Projection for Radiance Probes

The projection of a signal into a basis is an important and frequent problem in light transport simulation. For example, radiosity and precomputed radiance transfer rely on a representation of the light energy density in a scene in a suitable basis, and many off-line rendering systems employ bases to represent scattering functions. In fact, the ultimate goal of light transport simulation, the generation of digital images, is naturally formulated as the projection of

---

[41]Pinchon and Hoggan, "Rotation Matrices for Real Spherical Harmonics: General Rotations of Atomic Orbitals in Space-Fixed Axes".

[42]Kempski, *Extension of the Whittaker-Shannon Sampling Series Aided by Symbolic Computation*; Higgins, *Sampling Theory in Fourier and Signal Analysis: Foundations*.

[43]Freeden, Gervens, and Schreiner, *Constructive Approximation on the Sphere (With Applications to Geomathematics)*.

[44](Stern, "Classification Of Magnetic Shells"), see also the paper by James ("Transformation of Spherical Harmonics Under Change of Reference Frame").

[45]Gimbutas and Greengard, "A fast and stable method for rotating spherical harmonic expansions".
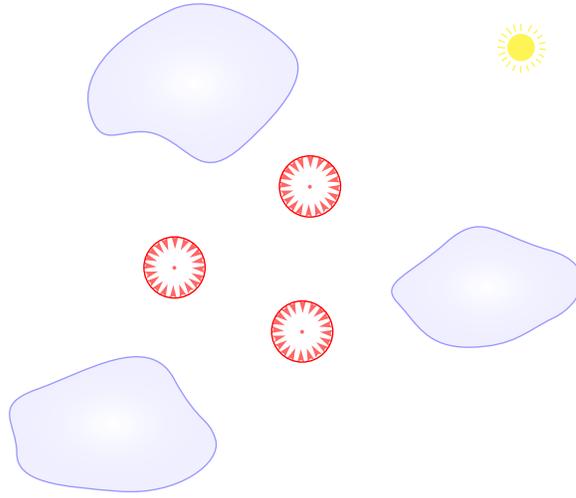
Figure 4.6: Radiance probes (red) represent the incidence light energy density $\ell_{q_i}(\omega)$ at a set of locations $q_i \in Q$ in a light transport scene in a suitable basis such as spherical harmonics.

the measured light energy density on the image plane into a finite dimensional basis.[46] In the following, we will use the results of Chapter 4.2.1.2 to determine the projection of the light energy density into a basis for the use as radiance probes, and we will employ the error analysis of Proposition 4.6 to understand and explain the superior performance of our technique compared to Monte Carlo and Quasi Monte Carlo integration, cf. Fig. 4.7.

Approximate Basis Projection   Radiance probes[47] represent incident monochromatic light energy density $\ell(q, \omega) \in S^2 Q$ at a set of locations $q_i \in Q$ in a light transport scene $(Q, n, \mathcal{M}, \rho)$, typically with homogeneous refractive index, cf. Fig. 4.6. Following the literature, we will employ spherical harmonics up to band L for the representation of $\ell_{q_i}(\omega) = \ell(q_i, \omega)$ in each fiber $S^2_{q_i} Q$. The approximation space of interest is hence $\mathcal{H}_L(S^2_{q_i} Q) = \mathcal{H}_{\leq L}(S^2_{q_i} Q)$, with the reproducing kernel being

$$k_L(\bar{\omega}, \omega) = \sum_{l=0}^{L} \sum_{m=-l}^{l} y_{lm}(\bar{\omega}) \, y_{lm}(\omega) = \sum_{l=0}^{L} \frac{2l+1}{4\pi} P_l(\bar{\omega} \cdot \omega), \qquad (4.44)$$

cf. Eq. 2.60. With the results from Chapter 4.2.1.2, the sought after spherical harmonics basis function coefficients $\beta_{lm}(q_i) = \langle \ell_{q_i}(\omega), y_{lm}(\omega) \rangle$ of the light

---

[46]From the point of view of our theory, the classical notion of a pixel as a "little square" corresponds to a characteristic basis, although much confusion on the subject exists in the literature, see for example (Fiume, *The Mathematical Structure of Raster Graphics*; Smith, *A Pixel Is Not a Little Square*; Blinn, "What Is a Pixel?").

[47]Radiance probes are sometimes also referred to as irradiance volumes or volumetric precomputed radiance transfer

energy density $\ell_i(\omega) = \ell_{q_i}(\omega)$ can be obtained by

$$\beta_{lm}(q_i) = \sum_{j=1}^{m} \ell_i(\lambda_j)\, r_{jk} \tag{4.45a}$$

where k is a linear index corresponding to (l, m), and the r_{jk} are the elements of the reconstruction matrix $\mathrm{R}_L(\Lambda)$ for a set of reproducing points Λ which are assumed to be independent of the probe location. In matrix-vector notion, the above equation takes the form

$$\beta(q_i) = \mathrm{R}_L(\Lambda)\, \ell_i(\Lambda) \tag{4.45b}$$

where $\beta(q_i) = (\beta_{0,0}(q_i), \dots, \beta_{L,L}(q_i))$. As in the previous section, next to using well distributed point sets Λ for the sphere which were obtained from low discrepancy sequences for the unit square, we also optimized the kernel locations λ_i using numerical optimization and with the condition number as quality measure. As shown in Table 4.1, oversampling together with optimization again yields nearly tight reproducing kernel bases with condition numbers very close to unity.

Although the desired approximation space is $\mathcal{H}_L$, there is no reason to believe that $\ell_i(\omega)$ will be contained in the space. Hence, as discussed in Chapter 4.2.1.2, employing Eq. 4.45 will incur error through the residual signal outside of $\mathcal{H}_L$. We will discuss this error in detail in the following.

Experimental Evaluation[48]   Experimental results for three test scenes, the classical Cornell box, the well known Sponza scene, and a room scene with teapots, are presented in Fig. 4.7.[49] Shown there is the L_2 error averaged over a large number of probe locations $q_i \in Q$ in the scene as a function of the number of samples. For comparison, we employed Monte Carlo and Quasi Monte Carlo integration, the latter one with the Halton sequence, to determine the integral of the product function $\ell(\omega)\, y_{lm}(\omega)$. The experimental results demonstrate that our technique (QA, opt.) outperforms Monte Carlo (MC) and Quasi Monte Carlo (QMC, Halton) integration, and it provides a significant advantage in particular when optimized kernel locations are employed. Unsurprisingly, Quasi Monte Carlo integration also clearly outperforms Monte Carlo integration with purely random sampling.[50]

In Fig. 4.7, top, it might seem surprising that Quasi Monte Carlo integration (QMC, opt.) with sampling locations optimized for our technique performs almost as well as our approximation approach. However, as shown in Table 4.1, with optimized sampling locations the condition number $\mathrm{cond}(\mathrm{K}_L(\Lambda))$ is very close to unity and the representation forms a nearly tight kernel frame.

---

[48]The experimental results were obtained using a modified version of the radiance probes implementation in the pbrt renderer (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*).

[49]Images of the scenes are available at http://www.pbrt.org/scenes.php.

[50]Keller, "Myths of Computer Graphics".
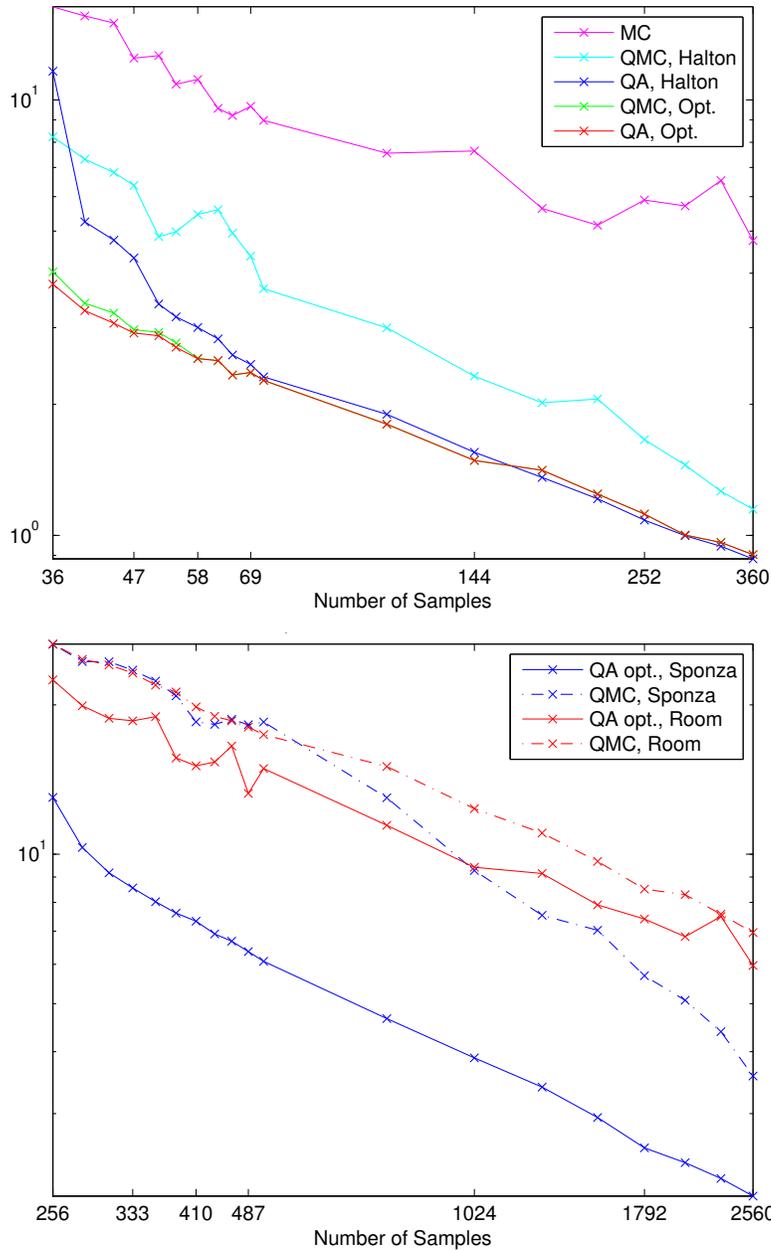
Figure 4.7: Average L_2 error for spherical harmonics basis projection for radiance probes; top, Cornell box scene for $\mathcal{H}_5$; bottom, Sponza and Room scene for $\mathcal{H}_{15}$. With optimized kernel locations, our technique (QA, opt.) considerably outperforms Monte Carlo (MC) and Quasi Monte Carlo (QMC, Halton) integration, which are the state of the art in the literature.

Oversampling	1.00	1.10	1.50	2.0	5.0	10.0
Halton	348.06	28.37	5.34	2.96	1.52	1.21
Halton, opt.	2.56	1.72	1.14	1.03	1.02	1.02

Table 4.1: Condition number of the reproducing matrix for $\mathcal{H}_{15}$ as a function of the oversampling rate for sampling sequences formed by the Halton sequence and for optimized points initialized from the sequence.

Hence, the reconstruction matrix $R_L(\Lambda) = K_L^{-1}(\Lambda)$ is well approximated by the transpose $K_L^{-1} \approx K_L^T$ and the computations become

$$\beta_{lm} = \sum_i y_{lm}(\lambda_i) \, \ell(\lambda_i) \tag{4.46}$$

which, up to a constant, is the standard Quasi Monte Carlo estimator for the product function $\ell(\omega) \, y_{lm}(\omega)$. This demonstrates the practical benefits facilitated by nearly tight reproducing kernel bases, and it allows for example to generate the integration weights on the fly during computations.

A comparison between the right hand side of Fig. 4.7 and Table 4.1 verifies the close correlation between the condition number of the reproducing matrix and the resulting error. Clearly visible is also that increasing the overcompletness of the kernel representation, and hence the number of samples used, leads to a characteristic decay in the error, well known from and paralleling those for Monte Carlo and Quasi Monte Carlo integration.

The implementation we employed for our experiments determines the sample values $\ell_i(\lambda_j)$ from a photon map using final gathering.[51] This is advantageous for our technique and Quasi Monte Carlo integration, which both assume regularity in the integrand. Nonetheless, with our technique we clearly outperform Quasi Monte Carlo integration since the space of functions of bounded variation is still too pessimistic for the signals one encounters.

Error Analysis By Proposition 4.6, the error of the basis projection in Eq. 4.45 for input signals not contained in $\mathcal{H}_L$ is bounded by

$$|\mathrm{err}_{lm}(f)| \leq \|\check{f}_L\| \|\gamma_k\| \tag{4.47}$$

where $\check{f}_L$ is the signal component not contained in $\mathcal{H}_L$, and the vector $\gamma_k = (\gamma_1^k, \ldots, \gamma_n^k)$ is formed by

$$\gamma_i^k = \sum_{j=1}^m y_i(\lambda_j) \, r_{kj}, \tag{4.48}$$

where i is a linear index corresponding to (l, m). The error formula in Eq. 4.47 poses the question if it can explain our experimental results. The quickly decaying error for fixed L in Fig. 4.7 suggests that the residual signal $\check{f}_L \notin \mathcal{H}_L$ has to be sufficiently small, and that $\|\gamma_k\|$ has to decay with increasing

---

[51]Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*, Chapter 17.3.
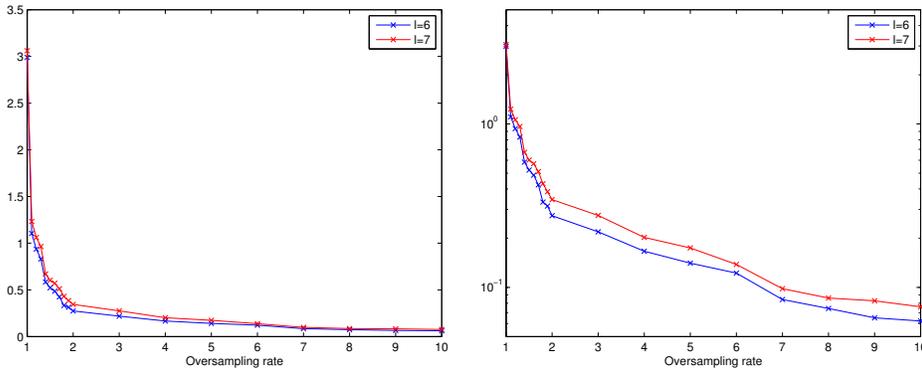
Figure 4.8: The effect of oversampling on $\|\boldsymbol{\gamma}_{lm}\|$ for approximate basis projection for the spherical harmonics space $\mathcal{H}_{\leq 5}$. Shown is $\|\boldsymbol{\gamma}_{lm}\|$ assuming the residual signal is contained in one band l. The left graph shows a linear plot, the right one a log plot.

oversampling rate. The norm of the signal components of $\ell_i(\omega)$ in each band $\mathcal{H}_l$, again averaged over a large number of radiance probe locations $q_i \in Q$, is shown in Fig. 4.9, top. It can be seen, that most of the energy is indeed contained in the lower bands, and hence in $\mathcal{H}_{\leq L}$. The behaviour of $\|\gamma_k\|$ as a function of the oversampling rate is shown in Fig. 4.8. The results demonstrate that the term decays rapidly with increasing overcompleteness, in particular for oversampling rates only slightly larger than unity. The error bound in Proposition 4.6 provides hence a useful characterization of the error that is observed in practice.

For the presented results we employed only function values $\ell_i(\lambda_j)$ of the incident light energy density $\ell_i(\omega)$, although we discussed previously in Chapter 4.2.1.1 and Chapter 4.2.1.2 that one could combine these with derivative values to obtain more information, and hence increase the sampling rate without increasing the number of samples, an idea which is particularly tantalizing for light transport simulation where more than 80% of the computation time is usually spend on sample generation.[52] However, as we mentioned in the foregoing, the derivative basis functions $\partial_\theta y_{lm}(\omega)$ and $\partial_\phi y_{lm}(\omega)$ are no longer normalized, and due to the increasingly oscillatory nature of spherical harmonics with increasing l the energy of the derivative in higher bands becomes more important,[53] see Fig. 4.9, bottom. Despite this increased energy in the high frequencies, our data suggests that derivative information can nonetheless be useful, and preliminary experimental results point in the same direction.

---

[52]Ibid.

[53]This is easily verified numerically, cf. www.dgp.toronto.edu/people/lessig/dissertation/LegendreDerivativeNormalized.nb. A formal proof is most likely possible using results from (Szegö, *Orthogonal Polynomials*; Freeden, Gervens, and Schreiner, *Constructive Approximation on the Sphere (With Applications to Geomathematics)*).
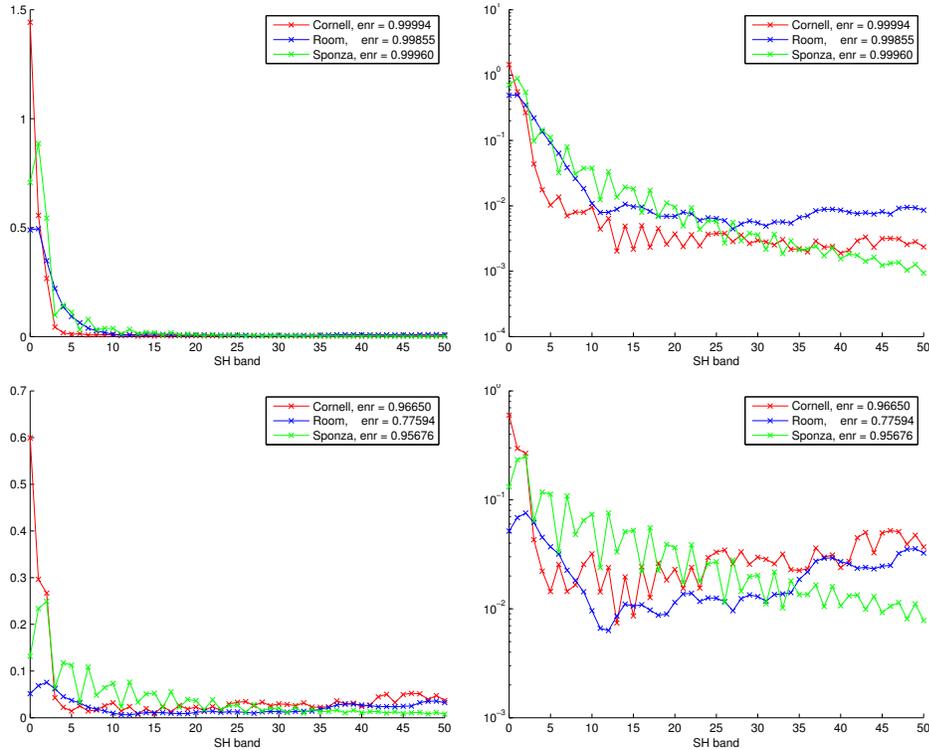
Figure 4.9: Top: Decay of the spherical harmonics coefficients for the three test scenes as linear (left) and log (right) plot. Together with the error analysis in Chapter 4.12, the decay explains the observed experimental results in Fig. 4.7. Bottom: Energy decay in the derivative signal. The behaviour suggests that derivative sampling could for many scenes be useful to improve performance. The energy in the first ten spherical harmonics bands is characterized by $\mathrm{enr} = \|f_{\leq 10}\|/\|f_{\leq 25}\|$.

4.2.4 Discussion

In this section, we employed biorthogonal and possibly overcomplete reproducing kernel bases to develop finitary point functionals—constructive computational techniques that employ only local information. These include for example sampling theorems, interpolation schemes, and integration rules, and our formulation provides an alternative perspective for classical techniques such as Monte Carlo integration, Lagrange interpolation, Gauss-Legendre quadrature, and the Shannon sampling theorem. We exemplified the potential of finitary point functionals with two applications, demonstrating the simplicity and the effectiveness of our approach.

In the following, we will relate finitary point functionals to work in the literature and discuss possible directions for future work. However, we have to caution that the connections that exist are too numerous to discuss all of them in detail.

Reproducing Kernel Hilbert Spaces in the Literature   As we have seen in the foregoing, many techniques in the literature can be expressed as finitary point functionals. Traditionally, however, none of them is formulated using this language, and instead a variety of approaches for sampling, interpolation, pointwise approximation, Monte Carlo integration, and quadrature rules are classically employed, with many of them being confined to peculiar function spaces and the real line. Nonetheless, reproducing kernels and reproducing kernel Hilbert spaces have recently received increased attention in various fields, such as signal processing,[54] approximation theory,[55] spline and data estimation,[56] integration theory,[57] machine learning,[58] computational harmonic analysis,[59] information based complexity theory,[60] and sampling theory.[61] Many of the recent results that employ reproducing kernel Hilbert spaces, however, are not widely known or acknowledged, and we are not aware of any discussion of classical quadrature rules or polynomial interpolation that employs reproducing kernels. Moreover, except for recent literature on sampling, none of the existing work employs bases or frames formed by reproducing kernel functions—and consequently none of it provides the common fabric for a wide range of techniques that is available with our formulation. In our opinion, this common perspective and the versatility of finitary point functionals, which can be defined for arbitrary finite function spaces over any domain, provide a vital advantage of reproducing kernel bases and finitary point functionals to work in the literature.[62]

---

[54]Unser, *"Sampling—50 Years After Shannon"*; Higgins, *Sampling Theory in Fourier and Signal Analysis: Foundations*.

[55]Reimer, *Multivariate Polynomial Approximation*; Cheney and Light, *A Course in Approximation Theory*.

[56]Wahba, *Spline Models for Observational Data*.

[57]Dick and Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*.

[58]Schölkopf and Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; Berlinet and Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*; Cucker and Smale, "On the Mathematical Foundations of Learning"; Poggio and Smale, "The Mathematics of Learning: Dealing with Data"; Smale and Zhou, "Shannon Sampling II: Connections to Learning Theory"; Cucker and Zhou, *Learning Theory: An Approximation Theory Viewpoint*.

[59]Saitoh, *Theory of Reproducing Kernels and its Applications*; Saitoh, *Integral Transforms, Reproducing Kernels and their Applications*; Coifman et al., "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Diffusion Maps"; Coifman et al., "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Multiscale Methods"; Coifman and Lafon, "Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions".

[60]Traub and Werschulz, *Complexity and Information*; Novak and Woźniakowski, *Tractability of Multivariate Problems: Standard Information for Functionals*.

[61]Nashed and Walter, "General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces"; Walter, "Wavelets and Sampling"; Higgins, *Sampling Theory in Fourier and Signal Analysis: Foundations*; Benedetto and Ferreira, *Modern Sampling Theory: Mathematics and Applications*.

[62]A similar but less complete conceptual framework based on reproducing kernels is provided in Schaback and Wendland, "Kernel techniques: From machine learning to meshless methods".

reproducing kernel	Dirac delta function, sampling function, point evaluation functional
dual kernel functions	reconstruction filter, interpolation function
reproducing points	sampling points, quadrature nodes, low discrepancy sequences, random samples
spanning point set	fundamental system, set of sampling, sampling set, spherical designs
reproducing matrix	Vandermonde matrix, discrete Fourier transform matrix
reconstruction matrix	interpolation matrix

Table 4.2: Sampling of the nomenclature for finitary point functionals found in the literature.

Spanning Point Sets and the Existence of Reproducing Kernel Bases
In the past, the only field which systematically considered bases and frames formed by reproducing kernel functions is sampling theory where the existence of sampling expansions in the form of Eq. 4.6 has been investigated, a question equivalent to the existence of spanning point sets which we left open in Chapter 4.1.[63] The results obtained there show that not all reproducing kernel Hilbert spaces admit sampling theorems, and hence there are spaces where no set of points is a spanning point set. A more precise form of the definition of finitary point functionals in Def. 4.5 should hence be restricted to reproducing kernel Hilbert spaces admitting sampling expansions.

For finite dimensional function spaces defined over continuous domains, the setting of most interest to us, it appears that spanning point sets should always exist. However, we do not have a definitive answer at the moment. Beyond existence, an important question is the characterization of conditions under which a set of locations is a spanning point set. A necessary requirement is that one sample location is in the support of every basis function, but it needs to be determined if this is also sufficient. For globally supported basis functions, such as the spherical harmonics we considered in our applications, almost any set of locations, up to a set of measure zero determined by the zero crossings of the functions, provides a spanning point set, paralleling the classical result for the Vandermonde matrix, cf. Chapter 4.2.2.3.[64]

An interesting open question is the existence and characterization of sampling

---

[63]See (Nashed and Walter, "General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces"; Mee, Nashed, and Seatzu, "Sampling Expansions and Interpolation in Unitarily Translation Invariant Reproducing Kernel Hilbert Spaces"; Han, Nashed, and Sun, "Sampling Expansions in Reproducing Kernel Hilbert and Banach Spaces"; Nashed and Sun, "Sampling and reconstruction of signals in a reproducing kernel subspace of $L_p(R^d)$") and references therein.

[64]The technique by Gimbutas and Greengard ("A fast and stable method for rotating spherical harmonic expansions") which we already discussed in Chapter 4.2.3.1 is an interesting example in this respect. There, all kernel locations are chosen on the equator and they hence do not provide a spanning point set. Gimbutas and Greengard therefore employ derivative sampling to reconstruct all basis function coefficients.

locations such that the reproducing kernel basis becomes an overcomplete, tight frame. The Shannon sampling theorem and characteristic bases provide examples of critically sampled reproducing kernel bases that are tight, but we are not aware of overcomplete examples. For the spherical harmonics spaces $\mathcal{H}_{\leq L}$ which we considered for our applications, it has been shown that no tight kernel basis can exist.[65] However, it seems unlikely to us that there are general obstructions to the existence, and we believe that an answer will depend on the reproducing kernel Hilbert space and the underlying domain. In our opinion, the most likely setting to encounter overcomplete, tight frames are as usual in highly symmetric spaces where the action of a group can be employed to generate sampling locations and to define suitable function spaces.

The difficulties associated with tight reproducing kernel bases motivated the concept of a nearly tight frame formed by reproducing functions. Such representations provide the same benefits as tight frames, for example the reproducing matrix does not have to be inverted and the reconstruction error is negligible, but for their construction it usually suffices to employ oversampling and to optimize the reproducing points. A better theoretical understanding of the construction of such representations and the performance that can be expected is however desirable, although one then might also encounter the aforementioned problems for tight reproducing kernel bases.[66]

Finite versus Infinite Dimensions We developed finitary point functionals as constructive mathematical techniques that readily admit a finite formulation—a formulation which is computationally practical.[67] Although we presumed in Chapter 4.2.1 often from the beginning a finite dimensional setting, the discussion there could easily be modified for infinite dimensional, separable reproducing kernel Hilbert spaces,[68] and Shannon's sampling theorem, which we considered in Chapter 4.2.2.4, is an example for a well known infinite dimensional finitary point functional.[69] Where we remained vague, and where our theory needs extension and refinement, is the connection to the infinite dimensional

---

[65]This is a consequence of the nonexistence of tight spherical designs (Bannai and Damerell, "Tight spherical designs, I"; Bannai and Damerell, "Tight Spherical Designs, II"), see also (Sloan and Womersley, "Extremal Systems of Points and Numerical Integration on the Sphere") for a discussion.

[66]For example, for the sphere S^2 it is known that the point distribution problem, which subsumes the questions of nearly optimal reproducing points on S^2, is very hard, and to this date only few concrete results have been obtained, see for example (Saff and Kuijlaars, "Distributing many points on a sphere"; Armentano, Beltrán, and Shub, "Minimizing the discrete logarithmic energy on the sphere: The role of random polynomials").

[67]Finitary point functionals hence shares the spirit of discrete exterior calculus with its "readily discretizable computational foundations", (Desbrun, Kanso, and Tong, "Discrete Differential Forms for Computational Modeling").

[68]We are not aware of results that ensure that a reproducing kernel Hilbert space is always separable.

[69]Much of the utility of Shannon's theorem, despite the problematic assumptions which are inherent in a bandlimited function over an infinite domain, arises from the translational invariance of the setting, which is the standard one in many engineering applications, see (Slepian, "On Bandwidth") for a discussion of these questions.

function spaces one usually encounters in applications and which are considered for example in mathematical physics. We believe that density arguments, which are the classical approach for this problem in constructive approximation,[70] are also an avenue for finitary point functionals, but for the moment it remains an open question how the gap is suitably bridged.

Error Analysis   We discussed two sources of error for finitary point functionals: reconstruction error, controlled by the condition number of the reproducing matrix, and the error that arises when the input signal is not contained in the space for which a technique was designed, aliasing error in classical nomenclature. We showed that the reconstruction error can be controlled through oversampling and by numerically optimizing the locations of the reproducing points, and for integration in finite reproducing kernel Hilbert spaces we presented in Proposition 4.7 a generalized Koksma-Hlawka inequality to bound this error. For signals not contained in an approximation space, we characterized the error in Proposition 4.6, and we showed that it is bounded by the signal's magnitude outside of the space, which would be the only error term for an optimal orthogonal projection, and a term that depends on the sampling locations. The characterization is constructive and it enables to determine well suited sampling locations which minimize the error. For classical quadrature rules, for example, it allows to justify the choice of the node locations, and for arbitrary function spaces and domains numerical optimization can be employed to determine sampling locations which minimize aliasing. Our error analysis also provides insight into the applicability of finitary point functionals and it allows to determine under which condition these are effective. This was demonstrated in Chapter 4.2.3.2 where Proposition 4.6 enabled us to explain the performance observed in practice. Proposition 4.6 was derived for the basis projection of a function from its values at a finite set of locations. It easily generalizes to other finitary point functionals since these employ a representation in a reproducing kernel basis. In the future, a mathematically more elegant and general formulation is however desirable.

Integration as Finitary Point Functional   A finitary point functional of central importance in many applications is integration. In Chapter 4.2.1.3 we employed reproducing kernel bases to determine a numerical integration rule that subsumes many techniques from the literature.[71] Classical quadrature rules are obtained from our formulation when orthogonal polynomials are considered, and Monte Carlo integration is recovered using characteristic bases with one sample in each partition. What is currently not subsumed by our formulation, at least when one insists on mathematical rigour, is Quasi Monte Carlo integration.

---

[70]See for example (DeVore and Lorentz, *Constructive approximation*; Pinkus, "Density in Approximation Theory").

[71]The earliest work that explicitly employed reproducing kernel Hilbert spaces to study integration that we are aware of is (Mysovskikh, "On the construction of cubature formulas with fewest nodes"), see (Cools and Sloan, "Minimal cubature formulae of trigonometric degree").

The spaces $\mathrm{BV}(X)$ of functions of bounded variation, which provide the setting for these integration rules, are only Banach spaces, cf. Example 2.14, and hence they do not admit a reproducing kernel in the sense of Chapter 2.2.3. However, using pairings as in Def. 2.29 instead of the Riesz representation theorem, the concept of a reproducing kernel has recently been extended to Banach spaces.[72] The dual spaces that then arise are isometrically isomorphic to measure spaces with bounded variation, which makes it tantalizing to assume that also classical Quasi Monte Carlo integration can be formulated using an ansatz analogous to finitary point functionals. Next to our formulation of numerical integration using reproducing kernel Hilbert spaces, these spaces also play a central role in recent work on integration theory, for example on low discrepancy sequences and the tractability of integration. In fact, for us this work provided important support for our thinking.[73]

In applied mathematics, an extensive literature on quadrature rules exist, and even those concerned with spherical harmonics spaces is too large to be surveyed here.[74] Similar to other fields, however, reproducing kernel Hilbert spaces play an increasingly prominent role in this community. For quadrature rules, also the numerical optimization of node locations has been considered, and the obtained results parallel our observations,[75] although the determinant, which

[72]Song, Zhang, and Hickernell, "Reproducing Kernel Banach Spaces with the ℓ_1 Norm"; Song and Zhang, "Reproducing Kernel Banach Spaces with the ℓ_1 Norm II: Error Analysis for Regularized Least Square Regression."; Zhang, Xu, and Zhang, "Reproducing Kernel Banach Spaces for Machine Learning"; Zhang and Zhang, "Frames, Riesz bases, and sampling expansions in Banach spaces via semi-inner products".

[73]See for example the recent monograph by Dick and Pillichshammer (*Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*). There, the similarity between $\mathrm{BV}([a,b])$ and the Sobolev space $H^1([a,b])$, cf. Example 2.13 and Example 2.14, is exploited to re-derive much of Quasi Monte Carlo theory using the reproducing kernel Hilbert space structure of $H^1([a,b])$, cf. Example 2.27. Interestingly, even in the integration theory community the numerical optimization of low discrepancy sequences has recently been considered (Dick et al., "Construction algorithms for polynomial lattice rules for multivariate integration"; Sloan, Kuo, and Joe, "On the Step-by-Step Construction of Quasi-Monte Carlo Integration Rules That Achieve Strong Tractability Error Bounds in Weighted Sobolev Spaces"; Sloan, Kuo, and Joe, "Constructing Randomly Shifted Lattice Rules in Weighted Sobolev Spaces"), although the work is probably best described as computer assisted proofs.

[74]The recent survey article (Hesse, Sloan, and Womersley, "Numerical Integration on the Sphere") is a good starting point, but see also (Xiao, Rokhlin, and Yarvin, "Prolate spheroidal wavefunctions, quadrature and interpolation"; Mhaskar, Narcowich, and Ward, "Spherical Marcinkiewicz-Zygmund Inequalities and Positive Quadrature"; Mhaskar, Narcowich, and Ward, "Representing and analyzing scattered data on spheres"; Sloan and Womersley, "Extremal Systems of Points and Numerical Integration on the Sphere"; Hesse and Sloan, "Worst-case errors in a Sobolev space setting for cubature over the sphere S^2"; Hesse and Sloan, "Cubature over the sphere S^2 x S^2 in Sobolev spaces of arbitrary order"; Keiner and Potts, "Fast evaluation of quadrature formulae on the sphere"; Hesse, "Complexity of numerical integration over spherical caps in a Sobolev space setting"; Gräf and Potts, "On the computation of spherical designs by a new optimization approach based on fast spherical Fourier transforms"; Gräf, Potts, and Steidl, *Quadrature rules, discrepancies and their relations to halftoning on the torus and the sphere*) and references therein.

[75]See for example (Reimer and Sündermann, "A Remez-type Algorithm for the Calculation of Extremal Fundamental Systems for Polynomial Spaces on the Sphere"; Reimer, *Multivariate Polynomial Approximation*; Sloan and Womersley, "Extremal Systems of Points and Numerical Integration on the Sphere").

was used as quality measure in this work, proved in preliminary experiments far less effective than the condition number used by us. Related to our formulation of integration are also fast transform algorithms, such as those for the Fourier series and spherical harmonics.[76] We believe that many of these transforms, including those for wavelets, are intimately related to reproducing kernels.

Arguably, the most interesting and most surprising reformulation of a classical technique in the literature as a finitary point functional is those for Monte Carlo integration. In contrast to traditional probabilistic results, this enables to precisely characterizes when the integration technique is accurate, and which error occurs when a signal does not lie in the finite characteristic function space implicitly employed. The practical limitation of the formulation is that a set of samples does not imply a partition, although the problem is more general since any set of samples leaves open the question which function space should be considered. An important motivation and influence for our finitary formulation of Monte Carlo integration was work in information based complexity theory.[77] There, it was shown that the classical setting of the technique, for example $L_2(X)$, is not computationally tractable: while one is guaranteed to converge ad infinitum, obtaining an acceptable error will in general also take an infinite amount of time. For example, for any function $f(x) \in L_2(X)$, its "evil sibling"

$$\bar{f}_\kappa(x) = f(x) + \kappa \prod_{i=1}^{n}(x - \lambda_i)^2$$

provides for any finite number of samples at locations λ_i an integration error which can be arbitrarily large for an appropriate value of κ.[78] One could argue, when $f(x)$ is of interest, then one will not accidentally encounter the "evil sibling" $\bar{f}_\kappa(x)$. However, the argument one is making then is that the functions of interest lie in a subset of $L_2(X)$ and thus have more regularity, or are "better behaved", than arbitrary finite energy signals—and it is this regularity which guarantees that one will not encounter $\bar{f}_\kappa(x)$ and hence ensures computational tractability. But then the regularity can also be exploited for more effective computations, and this is what leads to the better performance of Quasi Monte Carlo integration, which presumes the space $BV(X)$ of functions of bounded variation,[79] and of finitary point functionals, where finite reproducing kernel Hilbert spaces are employed. Indeed, the superior experimental results for our technique in Chapter 4.2.3.2 result from the local smoothness of the light energy density in each fiber, which can be exploited even with globally smooth spherical

---

[76]Driscoll and Healy, "Computing Fourier Transforms and Convolutions on the 2-Sphere"; Healy et al., "FFTs for the 2-Sphere-Improvements and Variations".

[77]For an introduction see (Traub and Werschulz, *Complexity and Information*), and for more technical discussions (Novak and Woźniakowski, *Tractability of Multivariate Problems: Linear Information*; Novak and Woźniakowski, *Tractability of Multivariate Problems: Standard Information for Functionals*).

[78]Traub and Werschulz, *Complexity and Information*.

[79]Dick and Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, p. 13.

harmonics and for which $\mathrm{BV}(S^2)$ is still too pessimistic as a setting. Ironically, it is in fact the empirical success of Monte Carlo integration in applications, and the convergence in finite time observed there, that provides the strongest evidence that the classical setting of the technique is too general.

An important argument for the use of Monte Carlo integration is often that the convergence rate is independent of the dimensionality of a problem and does not suffer from the so called "curse of dimensionality". However, recent results show that when Monte Carlo integration is employed for high dimensional problems, then an implicit smoothness is assumed,[80] and this is in fact necessary for Monte Carlo integration to be tractable in these settings.[81] We will return to the "curse of dimensionality" in more detail at the end of the next section in Chapter 4.3.3.

Compared to the literature, our ansatz to numerical integration provides the advantages of being constructive, versatile, and insightful. The effectiveness of our formulation stems from its finite dimensional setting which enables to obtain close to optimal node locations, low discrepancy sequences for arbitrary function spaces, using numerical optimization. Additionally, our integration rules employ weights, which can compensate for imperfect locations and provides an advantage over unweighted techniques such as Monte Carlo and Quasi Monte Carlo integration. Our experimental results and our error analysis demonstrate that even when idealized assumptions are not satisfied, we can still attain performance that considerably improves over the literature. The key to this robustness is the use of overcomplete representations which shields us against aliasing error from partial or incomplete information about the characteristics of the input signal. These advantages of our approach provide practical answers to the questions in the literature which we raised at the beginning of the chapter, cf. Page 402, and these will be collected at its closing in Chapter 4.5.

Applications To demonstrate the practical relevance of finitary point functionals, we considered two applications: the rotation of finite spherical harmonics expansions and the fiber-wise projection of the light energy density into a basis. The computational techniques usually employed for these problems are unrelated, spherical harmonics rotation is computed using matrices determined by recurrence schemes while Monte Carlo and Quasi Monte Carlo integration is employed for basis projection. Using finitary point functionals, we can employ variations of the same idea for both applications, and the same insights, for example that the use of overcomplete representations with optimized reproducing points is vital for performance, then apply for seemingly unrelated problems. Although it was no coincidence that spherical harmonics were employed in both examples, it was their analytic convenience that motivated our choice, and our formulation of finitary point functionals applies to arbitrary bases and function spaces. For example, for the light energy density it will be interesting to explore

---

[80]See (Donoho, "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality"; Bungartz and Griebel, "Sparse grids").

[81]This leads to the notion of a weighted function space where the magnitude of the signal component in higher dimensions decreases. We will return to such spaces in the following.

representations that are likely to correlate better with the signal properties
than globally smooth spherical harmonics.[82] Numerically, it would also be
interesting to explore more specialized algorithms to obtain dual frames, which
maybe be combined with our technique to optimize sampling points and then
should considerably improve the performance of sampling point optimization.[83]

Finitary Point Functionals for Light Transport Simulation   Most ap-
plications of light transport theory consider complex environments where only
a pointwise evaluation of the light energy density is possible. Classically, this
motivated the use of Monte Carlo and density estimation techniques. Our
finitary point functionals provide an alternative, functional analytic framework
to perform computations in this setting. In contrast to previous approaches,
the formulation is well defined with a finite number of samples, and it enables
to construct close to optimal techniques by numerical optimization. Addition-
ally, the experimental results indicate that the characteristics of the energy
density assure the effectiveness of our approach even when the input signal is
not perfectly contained in a finite approximation space. We therefore believe
that finitary point functional will provide an important tool for the numerical
simulation of light transport in the future.

The use of reproducing kernel bases to describe the transport of the light
energy density will be considered in detail in the next section. However, we
believe that finitary point functionals can prove useful for other aspects of the
problem that often do not receive the attention they deserve. For example, in
most rendering systems colors are represented using a discrete set of samples,
ever using specific wavelengths or by three color "channels". However, the
wavelength spectrum is a continuous signal, and hence the samples have to
represent basis function coefficients for a reproducing kernel basis. Respecting
this structure is important for example to ensure closure of the associated
function space under scattering and for a correct reconstruction.[84]

Another important application for finitary point functionals is the com-
putation of finite dimensional image representations, the projection of the
measurement of the light energy density over the image plane into a basis. In
the literature, the problem is still considered from the point of view of Shan-
non's theorem and assuming a representation of the input signal in the Fourier
domain.[85] Using samples obtained with ray tracing, the continuous image is
then determined with "reconstruction filters", chosen for example to optimize
the "visual appearance".[86] Reproducing kernel bases reveal that the choice of a

---

[82]See for example (Freeden, Gervens, and Schreiner, *Constructive Approximation on
the Sphere (With Applications to Geomathematics)*; Starck et al., "Wavelets, Ridgelets and
Curvelets on the Sphere").

[83](Gröchenig, "Acceleration of the frame algorithm"), see also (Mallat, *A Wavelet Tour of
Signal Processing: The Sparse Way*, Chapter 5.1).

[84]See (Peercy, "Linear Color Representations for Full Speed Spectral Rendering") for a
discussion of these aspects for general Hilbert space bases.

[85]See for example (Pharr and Humphreys, *Physically Based Rendering: From Theory to
Implementation*).

[86]Cook, "Stochastic Sampling in Computer Graphics"; Mitchell, "Generating Antialiased

reconstruction filter—a dual reproducing kernel basis function—is equivalent to a *global* choice of a function space, and unless the filter is the sinc-function, this is not the space of functions bandlimited in the Fourier domain. Additionally, our formulation shows that the reconstruction filters employed in the literature are questionable for the usual "pixel representations", or characteristic basis function in our parlance, that are employed in practice. A formulation of image sampling as a finitary point functional provides a rigorous mathematical foundation for the problem, and we believe that it will enable more efficient computational techniques, for example through sampling locations on the image plane which are designed for bases well suited for image representations, such as curvelets and bandlets.[87]

4.3 A Unified Formulation of Light Transport Simulation

In the literature, various approaches for the simulation of light energy density propagation in environments with scattering surfaces have been proposed. These can be categorized as finite element methods, such as radiosity, Monte Carlo algorithms, such as path tracing, and density estimation techniques, such as photon mapping, cf. Chapter 1.1.

Finite element methods and Galerkin projection are probably the most successful and most thoroughly understood techniques for the numerical treatment of continuous phenomena. Their applicability for light transport simulation, however, is limited since the complex environments of interest permit only a pointwise evaluation of the light energy density. Additionally, the high dimensionality of the energy density makes simplifying assumptions such as purely diffuse surfaces and distant illumination usually necessary. Monte Carlo methods, in contrast, employ naturally only point samples and do not require simplifying assumptions. However, as discussed in Chapter 4.2.4, these techniques cannot exploit regularity in the light energy density, and their usual formulation is questionable since only mathematical convergence but not computational tractability is assured. Additionally, even when the input signal is known, no precise information about the accuracy of Monte Carlo methods is available. Density estimation techniques are currently the most effective methods for light transport simulation and their performance is usually significantly better than those of Monte Carlo methods. However, currently no theoretical explanation for their efficacy exists, and the techniques suffer from the same limitation as Monte Carlo methods and for example only little understanding of their accuracy in finite settings is available.

Images at Low Sampling Densities"; Mitchell and Netravali, "Reconstruction filters in computer-graphics"; Mitchell, "Spectrally Optimal Sampling for Distribution Ray Tracing".

[87]Similar techniques have recently been proposed by Overbeck, Donner, and Ramamoorthi ("Adaptive Wavelet Rendering") and Hachisuka et al. (Hachisuka et al., "Multidimensional Adaptive Sampling and Reconstruction for Ray Tracing") but we believe finite reproducing kernel bases provides a more general and insightful theoretical framework which will yield more effective computational techniques.

With the lacking insight into computations with finite information, the setting unavoidable for practical calculations, one could argue that the benefits of Monte Carlo and density estimation techniques should be reconsidered, and in fact this has been suggested.[88] With the efficiency and accuracy provided by them in practice, however, we believe the argument should be turned around: the limitations should be interpreted as limitations of our theoretical understanding, and it is this which requires reconsideration. Maybe unsurprisingly, such an understanding can be gained with an alternative formulation of the techniques, and in fact not even a new formulation is needed. What was missing in the past, however, was a functional analytic elucidation of samples that allows to explain sampling based techniques using Galerkin projection. Reproducing kernel bases, as introduced in the foregoing, provide this functional analytic interpretation, and they consequently enable a formulation of light transport simulation where Monte Carlo and density estimation techniques are treated en par with finite element methods. Additionally, a functional analytic description provides insight into the working principles of the techniques, and it allows to understand their behaviour with a finite number of samples. Beyond the scientific insight that is provided, such a formulation also suggests alternative ways to develop more effective techniques for light transport simulation in the future.

To derive Monte Carlo and density estimation using Galerkin projection onto a reproducing kernel basis—an approach we call reproducing kernel Galerkin projection—we require a formulation of the light transport theory of Chapter 3 that more closely resembles those in the literature. Such a formulation will be presented in the next section. Afterwards, in Chapter 4.3.2 we will derive radiosity, distribution ray tracing, path tracing, and photon mapping, from the formulation. We conclude the section with a discussion of our approach and directions for future work in Chapter 4.3.3.

4.3.1 A Classical Formulation of Light Transport

In this section, we will recall our formulation of light transport from Chapter 3, and we will present it in a form which is similar to those usually employed in the computer graphics literature and as it was discussed in Chapter 1.1.1.

In Chapter 3.3.4, we derived an operator formulation for the transport of the phase space light energy density $\ell \in \mathrm{Den}(T^*Q) \cong \mathcal{F}(T^*Q)$ in a light transport scene $(Q, n, \mathcal{M}, \rho)$. There, it was shown that the outgoing steady state light energy density $\bar{\ell} \in \mathcal{H}(T^+\mathcal{M})$ on the scene surfaces $(\mathcal{M}, \rho)$ satisfies

$$\bar{\ell} = \ell_0 + \mathcal{T}^1 \ell_0 + \mathcal{T}^2 \ell_0 + \ldots = \sum_{i=0}^{\infty} \mathcal{T}^i \ell_0, \qquad (4.49a)$$

or equivalently as a balance equation

$$\bar{\ell} = \ell_0 + \mathcal{T}\bar{\ell}, \qquad (4.49b)$$

---

[88]Dahmen, *Private Communication*.
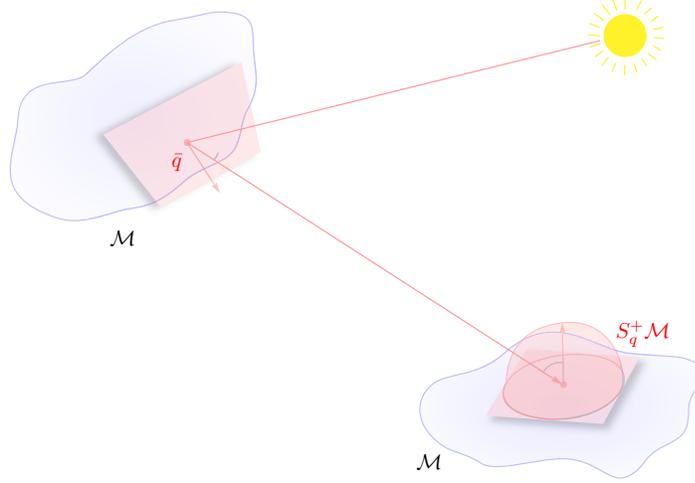
Figure 4.10: Geometry of the local description of light transport.

where $\ell_0 \in \mathcal{H}(T^+\mathcal{M})$ is an initial energy density, typically on the light sources. For this section, we will assume that the radiation is monochromatic and restricted to a suitably normalized value of the momentum $\omega = p \in S^*Q$, cf. Chapter 3.2.4. The steady state energy density is then

$$\bar{\ell} = \bar{\ell}(q, \omega) \in \mathcal{H}(S^+\mathcal{M}), \tag{4.50}$$

where the Hilbert space $\mathcal{H}(S^+\mathcal{M})$ is defined over the positive half-space $S^+\mathcal{M}$ of the cosphere bundle $S^*\mathcal{M}$ corresponding to outgoing directions, and the initial energy density is defined over the same space.[89] The operator $\mathcal{T}$ in Eq. 4.49 is the scattering transport operator

$$\mathcal{T} : \mathcal{H}(S^+\mathcal{M}) \to \mathcal{H}(S^+\mathcal{M}) \tag{4.51}$$

which by Eq. 4.50 also has to be restricted to the positive half-space $S^+\mathcal{M}$ of the cosphere bundle. By the definition of the scattering operator, Eq. 4.49b can locally at $q \in \mathcal{M}$ be written as

$$\bar{\ell}(q, \bar{\omega}) = \ell_0(q, \bar{\omega}) + \int_{H_q^+ Q} \left(\tilde{\eta}^*_{-t(z)} \, \bar{\ell} \right)(q, \omega) \, \tilde{\rho}_q(\omega, \bar{\omega}) \, d\omega \tag{4.52}$$

where $\tilde{\eta}_{-t(z)} = \tilde{\eta}_{t(z)}^{-1}$ is the flow on the cosphere bundle S^*Q to the previous surface, with the pullback $\tilde{\eta}^*_{-t(z)} \, \bar{\ell}$ relating outgoing to incoming light energy

---

[89]Note that $\ell(q, \omega)$ represents the light energy density parametrized in spherical coordinates and not the pre-radiance density $\tilde{\ell}(q, \omega)$. It is hence constant along trajectories in phase space, cf. Chapter 3.2.5.

density, see Def. 3.3 and Proposition 3.1. The integral kernel

$$\tilde{\rho}_q(\omega, \bar{\omega}) = \rho_q(\omega, \bar{\omega})\,(\vec{n}(q) \cdot \omega) \tag{4.53}$$

in Eq. 4.52 is the cosine-weighted surface scattering kernel in spherical coordinates, and, as is common in the literature, we will assume that it has support only in the fiber $S_q^2\mathcal{M}$ over q, cf. Fig. 4.10.[90] In slightly different notation, Eq. 4.52 is the "rendering equation" in Eq. 1.1 in Chapter 1.1.1.

Since every point $(q, \omega) \in S^*Q$ has a unique image on Q under the time evolution diffeomorphism $\tilde{\eta}_t$, we can employ the change of variables theorem to parametrize the light energy density over two surface points instead of surface location and direction. Computing the pullback of the relevant tensor basis functions of the light energy density yields for the change of variables[91]

$$d\omega(q) = \frac{\vec{n}(\bar{q}) \cdot \bar{\omega}_{\bar{q}}}{\|q - \bar{q}\|}\,dA(\bar{q}) \tag{4.54}$$

where $\bar{q} = \tilde{\eta}_{-t}(q)$ is the pre-image under the flow $\tilde{\eta}_{-t}$, the vector $\bar{\omega}_{\bar{q}} \in S_{\bar{q}}^2\mathcal{M}$ is the normalized direction vector from $\bar{q}$ to q, and $dA = \|\vec{n}\|\,du_1 du_2$ is the usual surface element on $\mathcal{M}$, cf. Example 2.114. With the area parametrization in Eq. 4.54, we obtain for Eq. 4.52 that

$$\bar{\ell}(q \to \bar{q}) = \ell_0(q \to \bar{q}) + \int_{\mathcal{M}} \bar{\ell}(\bar{q} \to q)\,\rho(\bar{q} \to q \to \bar{q})\,G(\bar{q} \to q)\,dA(\bar{q}) \tag{4.55}$$

where also an appropriate change of variables has been performed for the scattering double form.[92] The geometry term or geometry factor $G(q \to q) : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is given by

$$G(\bar{q} \to q) = V(\bar{q}, q)\frac{(\vec{n}(\bar{q}) \cdot \bar{\omega}_{\bar{q}})\,(\vec{n}(q) \cdot \omega_q)}{\|q - \bar{q}\|^2} \tag{4.56}$$

where $V(\bar{q}, q) : \mathcal{M} \times \mathcal{M} \to \{0, 1\}$ is the binary visibility function between $\bar{q}$ and q. Note that with the area formulation the scattering transport operator becomes

$$\mathcal{T} : \mathcal{H}(\mathcal{M} \times \mathcal{M}) \to \mathcal{H}(\mathcal{M} \times \mathcal{M}) \tag{4.57}$$

and it is defined over the tensor product space $\mathcal{M} \times \mathcal{M}$. Up to the term representing the emitted radiance, Eq. 4.52 and Eq. 4.55 also provide the local description for the iterates

$$\ell^k = \mathcal{T}^k \ell_0 = \mathcal{T}\ell^{k-1} \tag{4.58}$$

---

[90] In this section, we will forego with the technicalities that arise when the scattering kernel is treated more correctly as a differential double form. These considerations are irrelevant for the present discussion, and the interested reader is referred to Chapter 3.2.7 for the details.

[91] A rigorous derivation of the change of variables can be found at http://www.dgp.toronto.edu/people/lessig/dissertation/files/area_formulation.pdf.

[92] It appears that this change of variables for the scattering "function" is never considered in the literature.

in the series expansion in Eq. 4.49a, where ℓ^k is the light energy density after the kth "bounce".

In the following, we will employ the local formulation of the balance equation in Eq. 4.52 to derive radiosity and distribution ray tracing, and to obtain path tracing and photon mapping the iterates in Eq. 4.58 will be required.

4.3.2 Algorithms for Light Transport Simulation

In the following, we will derive classical algorithms for light transport simulations using Galerkin projection and following the general approach that was outlined in Chapter 2.2.4.3. The use of reproducing kernel bases will thereby enable us to also recover Monte Carlo and density estimation techniques using this approach. The existence of a tight or nearly tight reproducing kernel basis will tacitly be assumed, and we will otherwise remain formal, without detailing the function spaces we are working in. The validity of this treatment will be discussed in Chapter 4.3.3.

4.3.2.1 Radiosity

The radiosity algorithm[93] provides a solution to the light transport problem in purely diffuse environments when the scattering kernel is given by

$$\rho_q(\omega, \bar{\omega}) = \rho_q(\omega) \tag{4.59}$$

and does not depend on the outgoing direction $\bar{\omega}$. The scattered light energy density is then uniform in the fibers $S_q^2\mathcal{M}$, and hence also the steady state energy density satisfies $\bar{\ell}(q) = \bar{\ell}(q, \omega)$. As is well known in the literature, the radiosity algorithm can be obtained from the local light transport equation parametrized over area in Eq. 4.55 by performing Galerkin projection. The basis that has to be considered is a characteristic basis $\Xi(\mathcal{M}) = \{\chi_i\}_{i=1}^n$ over $\mathcal{M}$, cf. Chapter 2.2.5.1, where no angular dependence is required since the light energy density $\bar{\ell}(q)$ does not depend on direction. Without loss of generality, we will for the moment also consider a surface point $q \in \mathcal{M}$ that does not emit light.

With the characteristic basis $\Xi(\mathcal{M})$ and one $\lambda_i \in \mathcal{M}$ in the support of every basis function χ_i, the light energy density can be represented as

$$\bar{\ell}(\bar{q}) = \sum_{i=1}^n \bar{\ell}(\lambda_i) \, \chi_i(\bar{q}) \tag{4.60}$$

where the second variable over $\mathcal{M} \times \mathcal{M}$ has been dropped since the outgoing light energy is independent of direction. With this representation of $\bar{\ell}(\bar{q})$, Eq. 4.55

---

[93]The algorithm was introduced in (Goral et al., "Modeling the Interaction of Light between Diffuse Surfaces"; Nishita and Nakamae, "Calculation of Interreflections and Its Representation Method"), and Kajiya ("The Rendering Equation") showed that it is a rigorous approximation to the light transport problem. A thorough discussion of the technique can be found in (Cohen and Wallace, *Radiosity and Realistic Image Synthesis*).

can be written as

$$\bar{\ell}(q) = \int_{\mathcal{M}} \left(\sum_{i=1}^{n} \bar{\ell}(\lambda_i)\, \chi_i(\bar{q}) \right) \rho(q)\, G(\bar{q} \to q)\, dA(\bar{q}) \tag{4.61a}$$

Interchanging summation and integration yields

$$\bar{\ell}(q) = \rho(q) \sum_{i=1}^{n} \bar{\ell}(\lambda_i) \int_{\mathcal{M}} \chi_i(\bar{q})\, G(\bar{q} \to q)\, dA(\bar{q}) \tag{4.61b}$$

and by projecting the continuous outgoing light energy density also into $\Xi(\mathcal{M})$ we obtain

$$\langle \bar{\ell}(q), \chi_j(q) \rangle = \left\langle \rho(q) \sum_{i=1}^{n} \bar{\ell}(\lambda_i) \int_{\mathcal{M}} \chi_i(\bar{q})\, G(\bar{q} \to q)\, dA(\bar{q}),\, \chi_j(q) \right\rangle. \tag{4.61c}$$

Since the sample values $\bar{\ell}(\lambda_i)$ do not depend on q we have

$$\bar{\ell}(\lambda_j) = \rho_j \sum_{i=1}^{n} \bar{\ell}(\lambda_i) \left\langle \int_{\mathcal{M}} \chi_i(\bar{q})\, G(\bar{q} \to q)\, dA(\bar{q}),\, \chi_j(q) \right\rangle \tag{4.61d}$$

where on the left hand side we employed the reproducing property of the $\chi_j(q)$, and we also assumed that the diffuse scattering function is constant over the support of each basis function yielding scalar coefficients ρ_j. Eq. 4.61d suggests to expand the inner product on the right hand and to introduce symmetric form factors

$$F_{ji} = \int_{\mathcal{M}} \int_{\mathcal{M}} G(\bar{q} \to q)\, \chi_i(\bar{q})\, \chi_j(q)\, dA(\bar{q})\, dA(q) \tag{4.61e}$$

which describe the energy interchange between χ_i and χ_j. With the F_{ji}, Eq. 4.61d can be written as

$$\bar{\ell}(\lambda_j) = \rho_j \sum_{i=1}^{n} F_{ji}\, \bar{\ell}(\lambda_i). \tag{4.61f}$$

The form factors are usually arranged in a matrix $F(\Lambda)$ so that the above equation becomes a matrix-vector product. When the emitted energy density ℓ_0 is included, the equation reads

$$\bar{\ell}(\Lambda) = \ell_0(\Lambda) + \rho(\Lambda)\, F(\Lambda)\, \bar{\ell}(\Lambda) \tag{4.62}$$

where $\bar{\ell}(\Lambda)$ and $\ell_0(\Lambda)$ are the vectors containing the basis function coefficients $\bar{\ell}(\lambda_i)$ and $\ell_0(\lambda_i)$, respectively. Rearranging terms yields

$$\ell_0(\Lambda) = (I - \rho(\Lambda)\, F(\Lambda))\, \bar{\ell}(\Lambda) \tag{4.63}$$

where I is the $n \times n$ identity matrix. The steady state solution $\bar{\ell}(\Lambda)$ can hence be obtained using common numerical solvers for linear systems. This is the classical radiosity algorithm. The above derivation can be extended to more sophisticated basis functions,[94] and also to non-diffuse environments, although the five or six dimensional phase space which then has to be considered is typically too large to effectively employ the ansatz.[95]

4.3.2.2 Distribution Ray Tracing

Distribution ray tracing[96] was one of the earliest techniques for the realistic simulation of synthetic environments. The algorithm extends Whitted ray tracing[97] by averaging over multiple sample rays, and this allows to include effects such as soft shadow and motion blur.

   We will consider the example of computing the outgoing light energy density using distribution ray tracing. Hence, we are interested in a solution of Eq. 4.52 when only pointwise values of the incoming light energy density are employed and these are obtained by tracing rays. As in the previous section, we will assume that $q \in \mathcal{M}$ is on a surface that does not emit light. Let $\{k_i(z)\}_{i=1}^{m}$ be a tight or nearly tight reproducing kernel basis for $\mathcal{H}(S^+\mathcal{M})$ defined over a spanning point set $\Lambda = \{\lambda_i\}_{i=1}^{m}$ with $\lambda_i = (q_i, \omega_i) \in S^+\mathcal{M}$. With the kernel basis $\{k_i(z)\}_{i=1}^{m}$, the outgoing steady state light energy density $\bar{\ell}(\bar{q}, \omega) \in \mathcal{H}(S^+\mathcal{M})$ at the previous "bounce" can be represented by its values $\bar{\ell}(\lambda_i)$ at the λ_i as

$$\bar{\ell}(\bar{q}, \omega) = \bar{\ell}(z) = \sum_{i=1}^{m} \bar{\ell}(\lambda_i)\, k_i(z), \tag{4.64}$$

cf. Fig. 4.10. Using the linearity of the pullback, we hence obtain for the incoming light energy density that appears in Eq. 4.52 that

$$\left(\tilde{\eta}^*_{-t(z)}\, \bar{\ell}\right)(\bar{z}) = \left(\tilde{\eta}^*_{-t(z)} \left(\sum_{i=1}^{m} \bar{\ell}(\lambda_i)\, k_i(z)\right)\right)(\bar{z}) \tag{4.65a}$$

$$= \sum_{i=1}^{m} \bar{\ell}(\lambda_i) \left(\tilde{\eta}^*_{-t(z)}\, k_i(z)\right)(\bar{z}). \tag{4.65b}$$

When the transport by $\tilde{\eta}_t$ corresponds to a unitary operator, then we have by Proposition 4.2 that the functions

$$\bar{k}_i(z) = \tilde{\eta}^*_{-t(z)}\, k_i(z) = \mathcal{U}_{-t}\, k_i(z) \tag{4.66}$$

---

[94]See for example (Zatz, "Galerkin Radiosity: A Higher Order Solution Method for Global Illumination"; Gortler et al., "Wavelet Radiosity"; Schröder et al., "Wavelet Projections for Radiosity").

[95]Cf. (Christensen, "Hierarchical Techniques for Glossy Global Illumination"; Christensen et al., "Global Illumination of Glossy Environments Using Wavelets and Importance"; Christensen et al., "Clustering for Glossy Global Illumination").

[96]Cook, Porter, and Carpenter, "Distributed Ray Tracing".

[97](Whitted, "An Improved Illumination Model for Shaded Display"), and in parallel work (Kay, "Transparency, Refraction, and Ray Tracing for Computer Synthesized Images"), although one could argue that it is Euclid with his *Optica* who should be credited for the idea.

form again a tight reproducing kernel basis for all t. However, the surface transport operator as defined in Def. 3.3 is not unitary but only an isometry. Nonetheless, for the moment we will assume that the $\bar{k}_i(z)$ form a tight reproducing kernel basis, and we will come back to this question later. With the representation of the incoming light energy density from Eq. 4.65 and Eq.4.66, Galerkin projection of Eq. 4.52 is given by

$$\bar{\ell}(q,\bar{\omega}) = \int_{S_q^+ Q} \left(\sum_{i=1}^{m} \bar{\ell}(\lambda_i)\, \bar{k}_i(z) \right) \tilde{\rho}_q(\omega,\bar{\omega})\, d\bar{\omega} \tag{4.67a}$$

and interchanging summation and integration yields

$$\bar{\ell}(q,\bar{\omega}) = \sum_{i=1}^{m} \bar{\ell}(\lambda_i) \int_{S_q^+ Q} \bar{k}_i(z)\, \tilde{\rho}_q(\omega,\bar{\omega})\, d\omega. \tag{4.67b}$$

By exploiting the reproducing property of the $\bar{k}_i(z)$ we obtain

$$\bar{\ell}(q,\bar{\omega}) = \sum_{i=1}^{m_q} \bar{\ell}_q(\eta_{\text{-}t}(q,\omega_i))\, \tilde{\rho}_q(\omega_i,\bar{\omega}). \tag{4.67c}$$

where only the m_q sampling points $\lambda_i = (q_i,\omega_i)$ in the fiber $S_q^2 \mathcal{M}$ still contribute since the scattering kernel has support only at q. By Eq. 4.65, we also wrote $\bar{\ell}(\lambda_i)$ in Eq. 4.67c as

$$\bar{\ell}_q(\eta_{\text{-}t}(q,\omega_i)) = \bar{\ell}(\lambda_i) \tag{4.67d}$$

which is naturally interpreted as a point sample of the outgoing light energy density obtained by tracing a ray from (q,ω_i) along the diffeomorphism $\tilde{\eta}_{\text{-}t}$: $\mathbb{R} \times S^* \mathcal{M} \to S^* \mathcal{M}$ to the previous surface intersection at λ_i. Eq. 4.67c is discrete in the incoming direction ω but still continuous in $\bar{\omega}$. By projecting the equation into the reproducing kernel basis also in the outgoing direction for a basis function defined by a reproducing point $\lambda_j = (q_j,\omega_j)$ in the fiber $S_q^2 \mathcal{M}$, we obtain

$$\langle \bar{\ell}(q,\bar{\omega}),\, k_j(q,\bar{\omega}) \rangle = \left\langle \sum_{i=1}^{m_q} \bar{\ell}_q(\eta_{\text{-}t}(q,\omega_i))\, \tilde{\rho}_q(\omega_i,\bar{\omega}),\, k_j(q,\bar{\omega}) \right\rangle \tag{4.67e}$$

and again exploiting the reproducing property yields

$$\bar{\ell}(q,\bar{\omega}_j) = \sum_{i=1}^{m_q} \bar{\ell}_q(\eta_{\text{-}t}(q,\omega_i))\, \tilde{\rho}_q(\omega_i,\bar{\omega}_j). \tag{4.67f}$$

This is a modern formulation of distribution ray tracing where the outgoing light energy density $\bar{\ell}(q,\bar{\omega}_j)$ in direction $\bar{\omega}_j$ is determined by averaging over the incoming light from the last surface $\bar{\ell}_q(\eta_{\text{-}t}(q,\omega_i))$ from directions ω_i weighted by the scattering kernel $\tilde{\rho}_q(\omega_i,\bar{\omega}_j)$ evaluated at these directions, cf. Eq. 1.12.

With Chapter 4.2.1.3 in mind, Eq. 4.67f is also naturally interpreted as a Quasi Monte Carlo estimator for $\bar{\ell}(q, \bar{\omega}_j)$. Comparing the derivation in Eq. 4.67 to those in Eq. 2.53 for general bases, we see that the scalars $\tilde{\rho}_q(\omega_i, \bar{\omega}_j)$ are just the elements of the operator matrix from Def. 2.53 for the reproducing kernel basis, that is

$$\tilde{\rho}_q(\omega_i, \bar{\omega}_j) = \left\langle k_j(\bar{z}) \,\middle|\, \tilde{\rho}(\bar{z}, z) \,\middle|\, \bar{k}_i(z) \right\rangle. \tag{4.68}$$

In the following, we will briefly discuss the differences that arise when the reproducing kernel basis is no longer tight or nearly tight and we have to distinguish between the reproducing kernel basis functions $k_i(z)$ and their duals $\tilde{k}_i(z)$. This is also equivalent to no longer requiring that the pulled back basis functions $\bar{k}_i(z)$ in Eq. 4.66 form a tight reproducing kernel basis. For Eq. 4.67a, we then obtain

$$\bar{\ell}(q, \bar{\omega}) = \int_{S_q^+ Q} \left(\sum_{i=1}^{m} \bar{\ell}(\lambda_i) \, \tilde{k}_i(z) \right) \tilde{\rho}_q(\omega, \bar{\omega}) \, d\bar{\omega} \tag{4.69a}$$

where we did not introduce additional notation for the transported dual basis functions $\tilde{k}_i(z)$. Interchanging summation and integration then yields

$$\bar{\ell}(q, \bar{\omega}) = \sum_{i=1}^{m} \bar{\ell}(\lambda_i) \int_{S_q^+ Q} \tilde{k}_i(z) \, \tilde{\rho}_q(\omega, \bar{\omega}) \, d\bar{\omega}. \tag{4.69b}$$

For the dual basis functions $\tilde{k}_i(z)$ there is no reproducing property that can be exploited. However, analogous to the integration rule in Eq. 4.17, when we also consider the basis projection over the outgoing direction from Eq. 4.67e, we can introduce weights

$$w_{ji} = \langle \tilde{\rho}(z, \bar{z}_j), \tilde{k}_i \rangle = \left\langle k_j(\bar{z}) \,\middle|\, \tilde{\rho}(\bar{z}, z) \,\middle|\, \tilde{k}_i(z) \right\rangle \tag{4.69c}$$

which are again just the elements of the operator matrix. For Eq. 4.69b we then obtain

$$\bar{\ell}(q, \bar{\omega}_j) = \sum_{i=1}^{m_q} w_{ji} \, \bar{\ell}_q(\eta_{\text{-}t}(q, \omega_i)). \tag{4.69d}$$

Compared to Eq. 4.67f, this can be interpreted as a weighted quadrature rule to determine the outgoing light energy density $\bar{\ell}(q, \bar{\omega}_j)$ at the sample location $(q, \bar{\omega}_j)$. Importantly, Eq. 4.69d still only employs the point values $\bar{\ell}_q(\eta_{\text{-}t}(q, \omega_i))$ of the light energy density as input, and hence still admits a natural interpretation as a ray tracing method.

4.3.2.3 Path Tracing

Path tracing was introduced by Kajiya[98] and it is today the classical algorithm for the solution of the full light transport problem. The algorithm can be

---

[98]Kajiya, "The Rendering Equation".

considered as a Monte Carlo integration over the infinite dimensional space of light transport paths where one averages over the contribution made by all possible trajectories going from the emitters in a light transport scene to a receiving surface.[99]

With the area formulation, monochromatic light energy density $\ell \in \mathcal{H}(\mathcal{M} \times \mathcal{M})$ is defined over the tensor product space $\mathcal{M} \times \mathcal{M}$ and a point in the domain corresponds to the generalized ray $\bar{q} \to q$ defined by the phase space diffeomorphism $\tilde{\eta}_t$, see again Fig. 4.10. A tight or nearly tight reproducing kernel basis $\{k_i(\bar{q} \to q)\}_{i=1}^m$ for $\mathcal{H}(\mathcal{M} \times \mathcal{M})$ is then defined over a point set

$$\Lambda = \{\lambda_i\}_{i=1}^m = \{(\bar{q}_i, q_i)\}_{i=1}^m = \{\,\bar{q}_i \to q_i\,\}_{i=1}^m \tag{4.70}$$

where each sample location corresponds to a trajectory from one surface to the next. Given such a basis, the light energy density admits the representation

$$\ell(\bar{q} \to q) = \sum_{i=1}^m \ell(\lambda_i) k_i(\bar{q} \to q) = \sum_{i=1}^m \ell(\bar{q}_i \to q_i) k_i(\bar{q} \to q) \tag{4.71}$$

with $\ell(\bar{q}_i \to q_i)$ being the energy transported along the ray $\bar{q}_i \to q_i$, and where the geometry term, and hence also visibility, is implicitly already present in the basis function coefficients. Using Galerkin projection as in the previous sections but now for the local form of the iterate $\ell^{k+1} = \mathcal{T}\ell^k$, we have

$$\ell^{k+1}(q \to \bar{\bar{q}}) = \int_{\mathcal{M}} \left(\sum_{i=1}^m \ell^k(\lambda_i) k_i(\bar{q} \to \bar{\bar{q}}) \right) \rho(\bar{q} \to q \to \bar{\bar{q}}) \, G(\bar{q} \to q) \, dA(\bar{q}) \tag{4.72a}$$

Interchanging summation and integration yields

$$\ell^{k+1}(q \to \bar{\bar{q}}) = \sum_{i=1}^m \ell^k(\lambda_i) \int_{\mathcal{M}} k_i(\bar{q} \to \bar{\bar{q}}) \, \rho(\bar{q} \to q \to \bar{\bar{q}}) \, G(\bar{q} \to q) \, dA(\bar{q}) \tag{4.72b}$$

and with the reproducing property of the $k_i(\bar{q} \to \bar{\bar{q}})$ we obtain

$$\ell^{k+1}(q \to \bar{\bar{q}}) = \sum_{i=1}^m \ell^k(\lambda_i) \, \rho(\lambda_i, q \to \bar{\bar{q}}) \, G(\lambda_i) \tag{4.72c}$$

where summation is again only over the samples $\lambda_i = \bar{q}_i \to q_i$ where $q_i = q$, but for simplicity we will continue writing m. The left hand side is still defined over the continuous space of rays originating at q. Projecting Eq. 4.72c into the reproducing kernel basis for a basis function "anchored" at a reproducing point satisfying $\lambda_i = q \to \bar{\bar{q}}_i$ yields

$$\left\langle \ell^{k+1}(q \to \bar{\bar{q}}), k_j(\bar{q} \to \bar{\bar{q}}) \right\rangle = \left\langle \sum_{i=1}^m \ell^k(\lambda_i) \, \rho(\lambda_i, q \to \bar{\bar{q}}) \, G(\lambda_i), \, k_j(\bar{q} \to \bar{\bar{q}}) \right\rangle \tag{4.72d}$$

[99]For a discussion of the general theory of path integration from a computational point of view see (Traub and Werschulz, *Complexity and Information*, Chapter 5). Various applications are discussed in (Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*).

and exploiting the reproducing property we obtain

$$\ell^{k+1}(\lambda_j) = \sum_{i=1}^{m} \ell^k(\lambda_i)\,\rho(\lambda_i, \lambda_j)\,G(\lambda_i). \tag{4.72e}$$

Eq. 4.72e is analogous to our previous result in Eq. 4.67f for distribution ray tracing. However, this time we are considering the iterates of the scattering transport operator $\mathcal{T}$ term by term, which allows to write ℓ^k in the above equation in terms of ℓ^{k-1}. We then obtain

$$\ell^{k+1}(\lambda_{i_0}) = \sum_{i_1=1}^{m} \left(\sum_{i_2=1}^{m} \ell^{k-1}(\lambda_{i_2})\,\rho(\lambda_{i_2}, \lambda_{i_1})\,G(\lambda_{i_2}) \right) \rho(\lambda_{i_1}, \lambda_{i_0})\,G(\lambda_{i_1}). \tag{4.73a}$$

Continuing this expansion recursively yields

$$\ell^{k+1}(\lambda_{i_0}) = \sum_{i_1=1}^{m} \left(\sum_{i_2=1}^{m} \cdots \left(\sum_{i_k=1}^{m} \ell^0(\lambda_{i_k})\,\rho(\lambda_{i_k}, \lambda_{i_{k-1}})\,G(\lambda_{i_k}) \right) \cdots \right.$$
$$\left. \cdots \rho(\lambda_{i_2}, \lambda_{i_1})\,G(\lambda_{i_2}) \right) \rho(\lambda_{i_1}, \lambda_{i_0})\,G(\lambda_{i_1}) \tag{4.73b}$$

and by re-arranging the terms we obtain

$$\ell^{k+1}(\lambda_{i_0}) = \sum_{i_1=1}^{m} \cdots \sum_{i_k=1}^{m} \ell^0(\lambda_{i_k})\,\rho(\lambda_{i_k}, \lambda_{i_{k-1}})\,G(\lambda_{i_k}) \cdots$$
$$\cdots \rho(\lambda_{i_1}, \lambda_{i_0})\,G(\lambda_{i_1}). \tag{4.73c}$$

This suggest to define a path of length k as

$$\lambda_\alpha^k = (\lambda_{i_k}, \cdots, \lambda_{i_0}) \in \mathcal{M}^{k+1} \tag{4.73d}$$

where $\alpha = (i_1, \cdots, i_k)$ is a multi-index and the $(k+1)$-fold tensor product $\Lambda \times \cdots \times \Lambda \subset \mathcal{M}^{k+1}$ represents path space. By introducing the path contribution function $\pi : \mathcal{M}^{k+1} \to \mathbb{R}$ as

$$\pi(\lambda_\alpha^k) = \rho(\lambda_{i_k}, \lambda_{i_{k-1}})\,G(\lambda_{i_k}) \cdots \rho(\lambda_{i_1}, \lambda_{i_0})\,G(\lambda_{i_1}) \tag{4.73e}$$

we can write Eq. 4.73c as

$$\ell^{k+1}(\lambda_{i_0}) = \sum_{|\alpha|=k} \ell^0(\lambda_\alpha^k)\,\pi(\lambda_\alpha^k) \tag{4.73f}$$

where the summation is over all possible sequence α whose length $|\alpha|$ equals k. Eq. 4.73f determines ℓ^{k+1} as the average over all paths λ_α^k. The result directly extends to the infinite series in Eq. 4.49a that determines the steady state light energy density $\bar{\ell}$, with each term being of the form of Eq. 4.73f, and this recovers classical path tracing, cf. Eq. 1.13.

We will comment on an interpretation of Russian roulette and other techniques that are employed to truncate the infinite series in Eq. 4.3.3 to a finite sum within our framework in Chapter 4.3.3.

4.3.2.4 Photon Mapping

Photon mapping[100] is one of the most popular techniques for light transport simulation in applications. The algorithm is usually understood as a form of density estimation from a set of samples of the iterated light energy density ℓ^k. The samples are known as "photons" and represented in a data structure known as "photon map", which gives the technique its name.[101] Images are usually generated with final gathering which performs two iterates of path tracing and then employs the energy density ℓ^k estimated from the photon map instead of additional recursions.

From the previous section we have that the iterated light energy density ℓ^{k-1} can be reconstructed from sample values $\ell^{k-1}(\lambda_i)$, for example obtained with path tracing, by

$$\ell^{k-1}(q \to \bar{\bar{q}}) = \sum_{j=1}^{m} \ell^{k-1}(\lambda_j)\, k_j(q \to \bar{\bar{q}}). \tag{4.74}$$

This enables to employ Eq. 4.73a and instead of a recursive expansion to use the above representation. This yields

$$\ell^{k+1}(\lambda_{i_0}) = \sum_{i_1=1}^{m} \left(\sum_{i_2=1}^{m} \left(\sum_{j=1}^{m} \ell^{k-1}(\lambda_j)\, k_j(\lambda_{i_2}) \right) \tilde{\rho}(\lambda_{i_2}, \lambda_{i_1}) \right) \tilde{\rho}(\lambda_{i_1}, \lambda_{i_0}) \tag{4.75}$$

which is photon mapping with final gathering where $\ell^{k-1}(q \to \bar{\bar{q}})$ is reconstructed from samples $\ell^{k-1}(\lambda_j)$. In the literature, the reproducing kernel basis functions $k_j(\lambda_{i_2})$ are known as reconstruction filters, and one usually assumes that these have local support so that the inner most sum has to be performed only over a subset of the samples $\ell^{k-1}(\lambda_j)$. In practice, also "photons" from different iterates k are often employed simultaneously when the steady state light energy density is sought, and by Eq. 4.75 this is possible by exploiting linearity in the equation.

4.3.3 Discussion

In the previous sections, we demonstrated how radiosity, distribution ray tracing, path tracing, and photon mapping—algorithms classically considered as either being finite element, Monte Carlo, or density estimation techniques—can all be derived using Galerkin projection of the operator formulation of light transport.

---

[100](Jensen and Christensen, "Photon Maps in Bidirectional Monte Carlo Ray Tracing of Complex Objects"; Jensen, "Importance Driven Path Tracing using the Photon Map"; Jensen, *Realistic Image Synthesis using Photon Mapping*). For a modern discussion see also (Pharr and Humphreys, *Physically Based Rendering: From Theory to Implementation*, Chapter 15.6).

[101]We cannot refrain from once again emphasizing that the point samples of the continuous light energy density that are employed in "photon mapping" are completely and entirely unrelated to the concept of a photon in quantum electrodynamics, and more generally modern physics, despite the intuitive connection.

We employed tight or nearly tight reproducing kernel bases to encompass sampling based techniques in the ansatz, and the point values of the light energy density employed there were then naturally provided by the basis function coefficients of the representations.[102] Our formulation enables an understanding of techniques such as path tracing and photon mapping with a finite number of samples—the setting unavoidable for practical computations—and we believe that it provides the foundation to improve their efficacy in the future.

The derivation of the radiosity algorithm using Galerkin projection is by now standard in the literature.[103] An interesting question that should be considered, however, is how the collocation method, where the error is determined pointwise, can be formulated using reproducing kernels.[104] For precomputed radiance transfer, which is typically also considered as a variation of radiosity,[105] our formulation explains how the low dimensional approximations of the light energy density $\ell(q_i, \omega)$ at the scene locations $q_i \in \mathcal{M}$ relates to the continuous light energy density over all of $S^+\mathcal{M}$.

For the derivation of distribution ray tracing using Galerkin projection, we had to consider the transport of the outgoing light energy density. When the transport on phase space corresponds to a unitary operator, such as in the case of light transport in an environment without scattering surfaces, the closure of a reproducing kernel basis under the Hamiltonian flow is ensured by Proposition 4.2. This enables to determine time evolution of the energy density pointwise while retaining a rigorous description of the continuous signal, a result which we consider to be of principal importance for light transport simulation in environments with varying refractive index. Moreover, we believe that propagating point samples using a symplectic integrator, which locally preserves the structure of the Hamiltonian flow, will provide a globally structure preserving integrator for the light energy density.[106] Unfortunately, when scattering at surface is included, the surface transport operator is no longer unitary and the closure of a reproducing kernel basis under transport most likely only holds in special circumstances. These difficulties provided one of the motivations to also consider Galerkin projection with a reproducing kernel basis that is not tight. Although the resulting technique still only employs point values of the light energy density and admits an interpretation as a ray tracing approach, with the weights which then have to be introduced the simplicity and elegance of classical distribution ray tracing is lost. However, next to accommodating the transport of a reproducing kernel basis in between surfaces, relaxing the tightness requirement enables to consider function spaces and reproducing points that are better adapted to the characteristics of an environment. We

---

[102]Reproducing kernels have in the past rarely been used for Galerkin projection. See however (Li and Liu, *Meshfree Particle Methods*).

[103]See for example (Cohen and Wallace, *Radiosity and Realistic Image Synthesis*), or (Arvo, Torrance, and Smits, "A Framework for the Analysis of Error in Global Illumination Algorithms") for a more detailed derivation.

[104]See for example (Cohen and Wallace, *Radiosity and Realistic Image Synthesis*, Chapter 3.5).

[105]Lehtinen, "A Framework for Precomputed and Captured Light Transport".

[106]See our discussion on the subject in Chapter 3.3.5 and the references provided there.

458

believe that this will be vital for obtaining improved performance in applications, analogous to the improvements we were able to report in Chapter 4.2.3.2. While we did not consider non-tight reproducing kernel bases for other techniques such as path tracing, similarly weighted equations can be obtained there.

To obtain a functional analytic interpretation of path tracing, we employed Galerkin projection of the iterate $\ell^{k+1} = \mathcal{T}\ell^k$ for a tight or nearly tight reproducing kernel basis for $\mathcal{H}(\mathcal{M} \times \mathcal{M})$. Recursively expanding the resulting equation, conceptually similar to the derivation of the path integral formulation in classical light transport theory,[107] then yielded path tracing as an average over the paths between sampling points that arise through the expansion. Importantly, with the local regularity of the light energy density that has been observed in practice, which suggests that the signal lies close to a finite albeit probably rather high dimensional function space, our derivation provides an explanation of the tractability of path tracing in practical settings.[108] An interesting avenue for future work is to employ derivative values for the technique. This will enable to better amortize the high computational costs of obtaining samples, similar to the use of derivative values which we already discussed in Chapter 4.2. A question we did not consider so far is the finite termination of the infinite series which determines the steady state light energy density when the iterates $\ell^{k+1} = \mathcal{T}\ell^k$ are used. Classically, Russian roulette, a form of importance sampling over the space of path lengths, is employed to make sampling of longer and longer paths increasingly unlikely.[109] With our derivation in Chapter 4.2.2.1 and results in the modern literature on integration theory, such a decreasing probability corresponds to a weighted space where higher dimensions contribute less and less to a function.[110] For light transport, such a weighting is naturally provided by the scattering operator whose norm is strictly bounded by unity, cf. Proposition 3.3, and for any error bound $\epsilon > 0$ it is hence sufficient to consider only finite path lengths.

With our interpretation of path tracing, photon mapping with final gathering was obtained by terminating the recursive expansion of paths after the second step, and instead reconstructing the light energy density from "photon" samples

---

[107]See for example Veach's thesis ("Robust Monte Carlo Methods for Light Transport Simulation", Chapter 8).

[108]It should be kept in mind in this context that casting 64 rays per pixel for a 1024×768 image corresponds from out vantage point to a $50, 331, 648$ dimensional function space.

[109]In practice, this probabilistic termination is always combined with a fixed absolute path length.

[110]The idea has been introduced in (Sloan and Woźniakowski, "When Are Quasi-Monte Carlo Algorithms Efficient for High Dimensional Integrals?") and see for example (Dick and Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Chapter 12) for a recent treatment. It is also known that such weighted spaces are intimately related to reproducing kernel Hilbert spaces (Traub and Werschulz, *Complexity and Information*; Novak and Woźniakowski, *Tractability of Multivariate Problems: Standard Information for Functionals*). Some intuition for the concept of a weighted space can be obtained by considering Sobolev spaces defined using the Fourier transform, for example over the real line or over the sphere, cf. Eq. 2.11, where the contribution with increasing band is required to decrease, enforcing additional regularity in the signal.

using the reproducing kernel basis.[111] This provides a theoretically sound basis for the reconstruction, and it allows for example to determine when photon mapping is accurate with a finite number of samples or which error is introduced when only locally supported reconstruction filters are employed. Our rigorous interpretation of the reconstruction step also explains the observed effectiveness of photon mapping in comparison to path tracing, since reconstruction directly exploits the local regularity in the signal which for path tracing is not explicitly used at intermediate path lengths.

In the derivations in the foregoing sections, we exploited that the light energy density lies in the function spaces that were employed, since only then the reproducing property, which was crucial to obtain our results, is satisfied. For radiosity, this is explicitly enforced by the computation of the form factors F_{ji}, which corresponds to a projection onto the space $\Xi(\mathcal{M})$, and by choosing an input signal in the space. A vital difference of sampling based techniques is that no explicit projection takes place but that the light energy density is assumed to lie in the function space that is employed. For these techniques, hence an implicit Galerkin projection takes place, and this also explains the alternative interpretations of the algorithms that exist in the literature. Nonetheless, analogous to the bandpass filters employed in classical sampling theory to ensure that the assumptions of Shannon's theorem are satisfied, one can image a similar "filtering" for sampling based techniques for light transport simulation.

We believe that many other techniques from the literature, such as instant radiosity,[112] light cuts,[113] meshless radiosity,[114] and recent extension of photon mapping,[115] can also be obtained using reproducing kernel Galerkin projection. The details, however, remain to be developed. A technique for light transport simulation that is currently not easily subsumed in our formulation is Metropolis light transport.[116] It will be interesting to explore how our ideas have to be extended to also formulate Markov Chain Monte Carlo techniques using reproducing kernel bases.

Conceptually, for techniques that employ only local information the formulation in this section enables to determine nearly optimal sampling locations for light transport using numerical optimization. However, for computing such location and for using them we face the curse of dimensionality: the exponential dependence of the computational resources on the dimension of the domain.[117] With classical techniques, this makes problems in more than three dimensions

---

[111]Previously, density estimation techniques were formulated within a Monte Carlo framework by Veach ("Robust Monte Carlo Methods for Light Transport Simulation", Appendix 4.A). The connection between density estimation and reproducing kernels is apparently well known at least in parts of the literature (Berlinet and Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*).

[112]Keller, "Instant Radiosity".

[113]Walter et al., "Lightcuts: A Scalable Approach to Illumination"; Walter et al., "Multidimensional Lightcuts".

[114]Lehtinen, "A Framework for Precomputed and Captured Light Transport".

[115]Hachisuka, Ogaki, and Jensen, "Progressive Photon Mapping"; Hachisuka and Jensen, "Stochastic Progressive Photon Mapping".

[116]Veach and Guibas, "Metropolis Light Transport".

[117]See (Donoho, "High-Dimensional Data Analysis: The Curses and Blessings of Dimen-

computationally intractable, and appeared as a challenge for light transport simulation already in previous work on finite element methods. The curse of dimensionality can be avoided by increasing the smoothness of functions with increasing dimensionality. In practice, however, such assumptions are usually unrealistic. Nonetheless, more subtle forms of the idea provide a practical solution. For example, for multi-dimensional functions $f : \mathbb{R}^n \to \mathbb{R}$ for which the Fourier transform $F(\nabla f)$ of the gradient ∇f is in L_1, the n-term approximation converges with $O(n^{1/2})$ independent of the dimension,[118] and other spaces with similar convergence rates exist.[119] Generalizing these ideas leads to sparse grids and related methods,[120] which break the curse of dimensionality for example by assuming bounded mixed derivatives. Moreover, for suitable representations, such as wavelets for one dimensional signals and curvelets and bandlets for two dimensional ones, the convergence rates even hold when the functions of interest have singularities.[121] We believe that a better understanding of the natural function spaces of the light energy density and exploiting its local regularity will allow to break the curse of dimensionality also for light transport simulation.

Two crucial questions were disregarded in the derivations in the previous sections: which function spaces are we working in, and how can tight or nearly tight reproducing kernel bases be constructed for these spaces, if they exist at all? Firstly, the existence question should settled: does *any* function space with a tight reproducing kernel basis exist, which allows to concretely perform the derivations? This can be answered in the affirmative, and a characteristic basis provides a tight reproducing kernel basis which allows to carry out the programme, a result which should be hardly surprising with our foregoing interpretation of Monte Carlo integration in Chapter 4.2.2.1. Moreover, with the density of the characteristic basis in $L_2(X)$, it also allows to recover the known asymptotic results for Monte Carlo techniques and density estimation. For a partition that respects visibility, characteristic bases also satisfy our assumption for distribution ray tracing that the transported basis functions form again a tight reproducing kernel basis, and this closely resembles optimal representations for images which are also provided subdivision schemes adapted

---

sionality") and also (Bungartz and Griebel, "Sparse grids") for a detailed discussion of this issue.

[118] This was the first modern result where the curse of dimensionality was broken in a surprising way (Barron, "Universal approximation bounds for superpositions of a sigmoidal function"). The convergence rate is in this case "incidentally" exactly those of Monte Carlo integration. The spaces of functions $f : \mathbb{R}^n \to \mathbb{R}$ satisfying $F(\nabla f) \in L_1$ are also related to the Paley-Wiener spaces $\Omega_B(R^n)$ that provide the setting of Shannon's sampling theorem, cf. Chapter 4.2.2.4.

[119] See (Bungartz and Griebel, "Sparse grids", Section 2) for details. Orthant schemes, which provide one of the alternative examples, correspond apparently to Monte Carlo integration, although we are unfortunately not aware of a discussion of this results in the literature, beyond the passing comment on it by Donoho ("High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality") and by Bungartz and Griebel ("Sparse grids").

[120] See (ibid.) for an introduction and survey. Interestingly, one of the origins of sparse grids are quadrature rules and interpolation (Smolyak, "Quadrature and interpolation formulas for tensor products of certain classes of functions"), and this is another connection between our work and the literature which deserves much more attention.

[121] See for example (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*).

to singularities.[122] For future work, desirable is however a characterization of all function spaces and the tight or nearly tight reproducing kernel bases which permit the derivations. For tight frames, the existence question has already been discussed in Chapter 4.2.4, and we remain doubtful about both existence and an easy characterization. For nearly tight frames, we believe that for finite function spaces again overcomplete bases with numerically optimized reproducing points provide the desired representations, although, as we discussed before, the nontrivial structure and the high dimensionality of the domains $S^+\mathcal{M}$ and $\mathcal{M} \times \mathcal{M}$ make the construction probably still rather challenging.

4.4 The Function Spaces of Light Transport

What is the functional analytic setting of light transport theory? And which approximation spaces are suitable for light transport simulation? While we do not have definitive answers for these questions at the moment, we will in the following collect some partial results on a functional analytic characterization of light transport in a scene with scattering surfaces. We will begin with a concrete analysis for a restricted setting, and then discuss the general structure, to the extent we currently understand it.

4.4.1 The Effective Dimension of Light Transport[123]

Many applications require the solution of the light transport problem over a neighborhood on the scene manifold. In the following, we will study optimal linear approximations of the outgoing light energy density over such regions, and characterize the effective dimension that is required to attain an approximation error that is bounded by a given $\epsilon > 0$. We will show that under plausible assumptions the dimensionality is described by the spectrum of the spatio-spectral concentration problem on the sphere, and this will allow us to improve on existing estimates for the dimension using a more insightful and precise derivation. Moreover, our ansatz is constructive, and it provides a representation which attains the optimal linear approximation rate.

---

[122]For light transport, the unitary structure of the transport operator is lost because of occlusion effects, and this prevents the closure of a tight reproducing kernel basis under time evolution. Unsurprisingly, these functional analytic issues are similar to those arising in the representations of two-dimensional images where also anisotropic, spatially extended singularities caused by occlusion effects are the central source of difficulties. For images, adaptive triangulations provide optimal convergence rates (ibid., Chapter 9.3.2), which are in our parlance just characteristic bases with the edges adapted to the singularities of the problem. However, even in the two-dimensional case no efficient algorithm for the construction of such triangulations exist (ibid., p. 474).

[123]The work in this section appeared before in (Lessig and Fiume, "On the Effective Dimension of Light Transport"). The required mathematical background material can be found in Chapter 2.2.5.2 and Chapter 2.2.5.3.

4.4.1.1 Locally Coherent Light Transport

Let $(Q, n, \mathcal{M}, \rho)$ be a light transport scene with homogeneous refractive index n. We are interested in scattering transport operators $\mathcal{T} : \mathcal{H}(S^+\mathcal{M}) \to \mathcal{H}(S^+\mathcal{M})$ that are locally radially symmetric around an axis $\vec{c}(q) \in S_q^-\mathcal{M}$, and whose action determines the monochromatic steady steady state light energy $\bar{\ell}(q) \in \mathcal{H}(S^+\mathcal{M})$ by

$$\bar{\ell}(q) = \int_{S_q^2\mathcal{M}} \bar{\ell}(\tilde{\eta}_{-t}(q, \omega))(\omega)\, T(\vec{c}(q) \cdot \omega, \bar{\omega})\, d\omega, \tag{4.76}$$

cf. Chapter 4.3.1. The incoming light energy density $\bar{\ell}(\tilde{\eta}_{-t}(q, \omega))(\omega)$ is assumed to be independent of the locations $q \in \mathcal{M}$, which is a valid approximation when the outgoing energy density is distant and far enough away from the region of interest,[124] and we will not be concerned with the transport by $\tilde{\eta}_t$ in this section and simply employ the incoming light energy density $e(\omega) = \bar{\ell}(\tilde{\eta}_{-t}(q, \omega))(\omega) \in S^-\mathcal{M}$ in the following, in which case Eq. 4.76 becomes

$$\bar{\ell}(q) = \int_{S_q^2\mathcal{M}} e(\omega)\, T(\vec{c}(q) \cdot \omega, \bar{\omega})\, d\omega. \tag{4.77}$$

Assuming $\mathcal{T}$ is bandlimited in the spherical harmonics domain, we study its optimal linear k-term approximations $\tilde{\mathcal{T}}$ such that the approximation error

$$\text{err}_k = \|\ell - \tilde{\ell}\|_{\mathcal{U}'}^2 = \|\mathcal{T}\ell - \tilde{\mathcal{T}}\ell\|_{\mathcal{U}'}^2 \tag{4.78}$$

of the outgoing light energy density $\bar{\ell}$ over a convex neighborhood $\mathcal{U}'$ on the scene manifold $\mathcal{M}$ is minimized, cf. Fig. 4.11. We show that for sufficiently large k, with the original size m of the problem typically much larger than k, the error in Eq. 4.78 becomes vanishingly small. The number k will then be referred to as the **effective dimension of light transport**, and the ratio k/m provides a measure for the local variation of light transport in the neighborhood $\mathcal{U}'$ on the scene manifold.

The foregoing assumptions on the scattering transport operator $\mathcal{T}$, local symmetry and bounded frequency in the spherical harmonics domain, are satisfied for example for a diffuse shading kernel

$$T(\vec{n}(q) \cdot \omega) = \rho(q)\,(\vec{n}(q) \cdot \omega) \tag{4.79}$$

where the axis of symmetry is the local normal $\vec{n}(q)$ and the bandlimit is $L = 2$.[125] Specular transport can be studied with the formulation in Eq. 4.77 when only a single outgoing direction $\tilde{\omega}$ is of interest, for example the direction to a camera, and partial occlusion will be analyzed by assuming that the incident light energy density is radially symmetric around a central axis.

---

[124]This is *the* central assumption in precomputed radiance transfer, cf. (Lehtinen et al., "A Meshless Hierarchical Representation for Light Transport").

[125]Ramamoorthi and Hanrahan, "Frequency Space Environment Map Rendering".
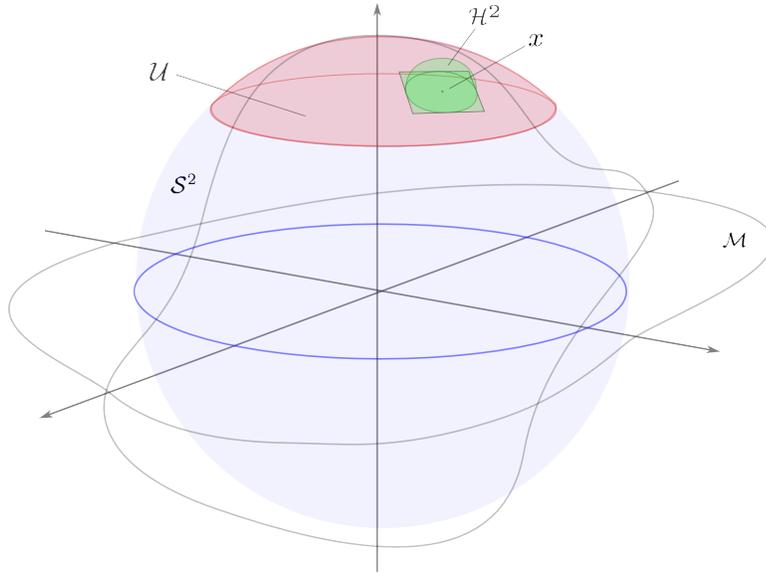
Figure 4.11: We study the effective dimension of the outgoing light energy density $\ell(q)$ over a region $\mathcal{U}' \subset \mathcal{M}$ when the scene manifold $\mathcal{M}$ can be approximated locally by a subset of the sphere $U \subset S_{\mathcal{M}}^2$, and the radially symmetric transport operator $\mathcal{T}$ acts as a local integral operator over the hemisphere $S_q^+\mathcal{M}$ at $q \in \mathcal{M}$, mapping distant incoming to outgoing light energy density.

The limited local variation of light transport has been exploited in precomputed radiance transfer for some time.[126] A theoretical analysis was recently undertaken by Mahajan and co-workers.[127]. There, the problem was considered in two dimensional "flatland", and results were obtained by estimating the spectrum of a discretized transport operator using a variation of Szegö's eigenvalue theorem. The extension to three dimensions was sketched by these authors. Using the same assumptions as Mahajan and co-workers, we study the effective dimension of light transport by reducing Eq. 4.78 to the spatio-spectral concentration problem on the sphere. This yields a derivation which provides greater insight, does not require a discretization, is directly applicable for the three-dimensional setting of interest, and moreover quantitatively improves upon their results.

4.4.1.2 Analysis of the Effective Dimension

To analyze the effective dimension of light transport, consider Eq. 4.77 and let the scattering transport operator $\mathcal{T}$ be L-bandlimited in the spherical harmonics domain. Furthermore, assume that $\vec{c}(q)$ is the local normal $\vec{n}(q)$; we will discuss

---

[126]Sloan et al., "Clustered Principal Components for Precomputed Radiance Transfer"; Liu et al., "All-Frequency Precomputed Radiance Transfer for Glossy Objects".

[127]Mahajan et al., "A Theory of Locally Low Dimensional Light Transport".

generalizations in the sequel. By its radial symmetry and the bounded frequency, $T(\vec{n}(q) \cdot \omega)$ is naturally represented using zonal harmonics $P_l(\vec{n}(q) \cdot \omega)$ centered at $\vec{n}(q)$,

$$T(\vec{n}(q) \cdot \omega) = \sum_{l=0}^{L} t_l \, P_l(\vec{n}(q) \cdot \omega) = \sum_{l=0}^{L} t_l \sum_{m=-l}^{l} y_{lm}(\vec{n}(q)) \, y_{lm}(\omega) \qquad (4.80)$$

where the right hand side follows from the spherical harmonics addition theorem, cf. Eq. 2.59. Expanding the incident radiance $e(\omega)$ in spherical harmonics yields

$$e(\omega) = \sum_{l'=0}^{L} \sum_{m'=-l'}^{l'} e_{l'm'} \, y_{l'm'}(\omega) \qquad (4.81)$$

and where only the expansion up to band L has to be considered since the residual signal will be in the kernel of $\mathcal{T}$. With the domain of integration extended to S^2, the domain of orthonormality of the y_{lm}, we can rewrite Eq. 4.77 in a global coordinate frame as

$$\bar{\ell}(q) = \int_{S^2} \left(\sum_{l'=0}^{L} \sum_{m'=-l'}^{l'} e_{l'm'} \, y_{l'm'}(\omega) \sum_{l=0}^{L} t_l \, P_l(\vec{n}(q) \cdot \omega) \right) d\omega. \qquad (4.82a)$$

Exploiting linearity, we obtain

$$\bar{\ell}(q) = \sum_{l'=0}^{L} \sum_{m'=-l'}^{l'} e_{l'm'} \sum_{l=0}^{L} \sum_{m=-l}^{l} t_l \, y_{lm}(\vec{n}(q)) \int_{S^2} y_{l'm'}(\omega) \, y_{lm}(\omega) \, d\omega \quad (4.82b)$$

and by the orthonormality of the spherical harmonics this yields

$$\bar{\ell}(q) = \sum_{l=0}^{L} \sum_{m=-l}^{l} e_{lm} \, t_l \, y_{lm}(\vec{n}(q)). \qquad (4.82c)$$

For a convex region $\mathcal{U}'$ on $\mathcal{M}$ which is well approximated by a subset $U \subset S_{\mathcal{M}}^2$ of the sphere $S_{\mathcal{M}}^2$, see Fig. 4.11, the local normal $\vec{n}(x)$ coincides with the direction vector $\bar{\omega}$ of $S_{\mathcal{M}}^2$. Eq. 4.82c can then be written as

$$\bar{\ell}(q) \approx \ell(\bar{\omega}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \bar{\ell}_{lm} \, t_l \, y_{lm}(\bar{\omega}) \qquad (4.83a)$$

and defining coefficients b_{lm} as $\bar{\ell}_{lm} \, t_l$ we obtain

$$\ell(\bar{\omega}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} b_{lm} \, y_{lm}(\bar{\omega}). \qquad (4.83b)$$

By Eq. 4.83, the outgoing energy density $\bar{\ell}(q)$ defined over $\mathcal{M}$ is approximated by the L-bandlimited signal $\bar{\ell}(\bar{\omega})$ defined over $S_{\mathcal{M}}^2$.
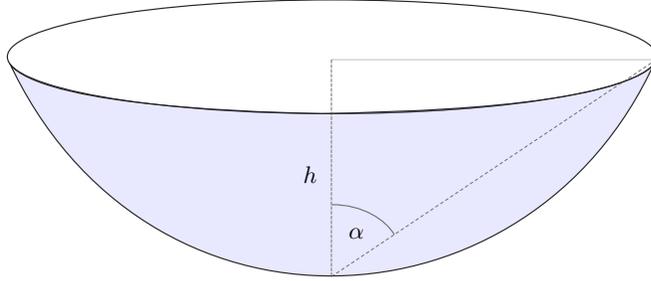
Figure 4.12: To study partial occlusion, we consider the interior of a hemispherical cap and a neighborhood U (blue) centered at the pole.

We will now turn to the study of an optimal linear k-term approximation of Eq. 4.83. Let $\{\varphi_i\}_{i=1}^M$ with $M = (L+1)^2$ be an arbitrary basis for $\mathcal{H}_L(S^2_{\mathcal{M}})$. A linear k-term approximation of Eq. 4.83 is obtained when $k \leq m$ basis functions $\{\varphi_i(\bar{\omega})\}_{i \in \mathcal{I}}$ are employed to represent $\bar{\ell}(\bar{\omega})$ and the index set $\mathcal{I}$ is chosen irrespective of the signal. Assuming the basis functions φ_i are orthogonal over U, the approximation error for the outgoing light energy density $\bar{\ell}(\bar{\omega})$ over $U \subset S^2_{\mathcal{M}}$ in Eq. 4.78 can be written as

$$\mathrm{err}_k = \|\bar{\ell}(\bar{\omega}) - \tilde{\ell}(\bar{\omega})\|_U = \sum_{i=K+1}^M \bar{b}_i^2 \langle \varphi_i, \varphi_i \rangle_U = \sum_{i=K+1}^M \bar{b}_i^2 \|\varphi_i\|_U^2. \qquad (4.84)$$

For arbitrary $\bar{\ell}$ and $\mathcal{T}$, and hence arbitrary $\bar{b}_i^2$, the error is thus minimized by basis functions $\varphi_i(\bar{\omega})$ whose squared norm $\|\varphi_i\|_U^2$ is minimal over U. By construction, however, the $m - k$ bandlimited functions whose norm is minimal over the region $U \subset S^2$ are the $m - k$ Slepian functions which are associated with the smallest eigenvalues of the spatio-spectral concentration problem. The optimal k-term approximation for $\bar{\ell}(\bar{\omega})$ is hence provided by the first k Slepian functions, and these naturally satisfy our previous orthogonality assumption. The sought estimate for the effective dimension of the locally symmetric scattering transport operator $\mathcal{T}$ in a neighborhood $\mathcal{U}' \subset \mathcal{M}$ is thus provided by

$$N(\epsilon) = \frac{C}{4\pi} + \log\left(\frac{1-\epsilon}{\epsilon}\right) B(\partial U) \log(C) + o(\log C) \qquad (4.85)$$

where $B(\partial U)$ is a function which depends on the boundary ∂U of the region of concentration and $C = (L+1)^2 A(U)$, cf. Chapter 2.2.5.3. For a sufficiently large region of spatio-spectral concentration, the effective dimension is hence well approximated by the Shannon number N, the first term of Eq. 4.85, and otherwise the generalized Shannon number N_g, given by the first two terms of Eq. 4.85, provides a tight characterization, cf. Fig. 2.8.

To study the effective dimension of light transport in partially occluded environments, we will assume that the incoming light energy density $\bar{\ell}(\omega) =$

$\ell(h \cdot \omega) \in S^- \mathcal{M}$ is radially symmetric with axis $h \in S^2$, cf. Fig. 4.12.[128] In the hemisphere above the origin, the incoming light energy density is then symmetric around the up axis and we assume that the neighborhood $\mathcal{U}$ is sufficiently small so that deviations from this assumption are negligible for all $q \in \mathcal{U}$. With a derivation analogous to those which led to Eq. 4.82c, one then obtains

$$B(x) = \sum_{l=0}^{L} e_l \, t_l \, P_l(h \cdot \vec{n}(x)). \approx \sum_{l=0}^{L} e_l \, t_l \, P_l(h \cdot \bar{\omega}). \tag{4.86}$$

where we again identified $\vec{n}(x)$ and $\bar{\omega} \in S^2_{\mathcal{M}}$. It follows that the best k-term approximation for $\bar{\ell}(\bar{\omega})$ over all of U is given by the L-bandlimited and radially symmetric functions which are optimally concentrated in U. Slepian function hence once again provide the optimal representation. If we furthermore assume that U is a spherical cap of co-latitude Θ, then by the radial symmetry of the $P_l(h \cdot \bar{\omega})$ in Eq. 4.86 the spatio-spectral region of concentration is $C = 4\,(L{+}1)\,\Theta$. For radial symmetric lighting and a region of concentration which is a spherical cap the dimensionality hence depends directly on the co-latitude Θ instead of the area $A(U)$.

Experimental results for the effective dimension of light transport are presented in Fig. 4.13, and clearly visible is the importance of the second term of the generalized Shannon number to accurately estimate the dimensionality. Differences between the theoretical and empirical spectra arise from magnitude variations of the basis function coefficients for real transport operators which we do not model.

4.4.1.3 Discussion

In this section, we showed that the effective dimension of a bandlimited and radially symmetric transport operator in a local neighborhood on the scene manifold can be studied using spatio-spectral concentration theory. An estimate for the dimensionality is given by the Shannon number N, and a more refined analysis leads to the generalized Shannon number N_g. This demonstrates that there exists both a linear and a logarithmic dependence on the region of concentration. The logarithmic term was not obtained in previous work,[129] although it is of particular importance for the settings considered in computer graphics where very low bandlimits L are employed. For radially symmetric lighting and a neighborhood U which is a spherical cap, a setting which allows one to study partial occlusion, we showed that the spatio-spectral region of concentration depends linearly on the co-latitude Θ of the spherical cap instead of the area $A(U)$, paralleling the result by Mahajan et al.[130] In addition to the refined estimates for the effective dimension, the results available by using

---

[128]The geometric setup consider by us is from (Ramamoorthi, Koudelka, and Belhumeur, "A Fourier Theory for Cast Shadows") for the same problem.

[129]Mahajan et al., "A Theory of Locally Low Dimensional Light Transport".

[130]Ibid.
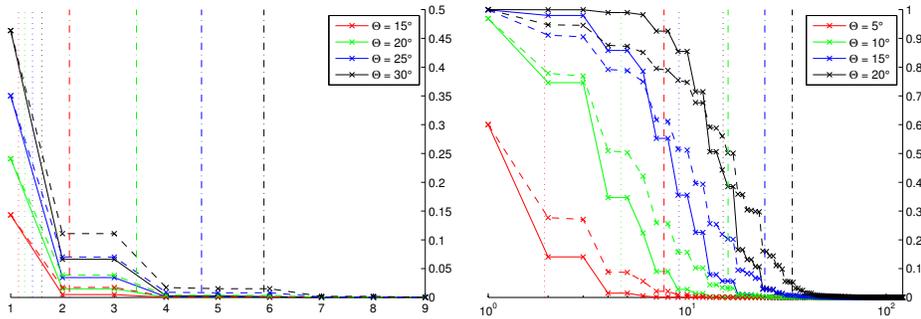
Figure 4.13: Spectrum of the spatio-spectral concentration problem (full) and for empirical transport operators (dashed, obtained using singular value decomposition) for spherical caps with $\theta \leq \Theta$. Left, diffuse transport $T(\vec{n}(q), \omega) = (\vec{n}(q) \cdot \omega)$ for $L = 2$; right, Phong transport $T(\mathrm{r}(q), \omega) = (\mathrm{r}(q) \cdot \omega)^{128}$ for $L = 10$, $\mathrm{r}(q)$ is the local reflection direction. The eigenvalue index is shown on the X axis and the magnitude $|\lambda_i|$ on the Y axis. Shown are also the Shannon number N (dotted) and the generalized Shannon number N_g (dash-dot), the latter one obtained with $B(\partial U) = \lg\left((L+1)^2 |\partial U|^2\right)/\lg(2\pi)$. Differences in the spectra between the predictions by spatio-spectral concentration theory and the singular value decomposition arise from the non-uniform energy distribution for the empirical transport operators across bands.

spatio-spectral concentration theory are more general than those obtained previously, for example they hold for arbitrarily shaped regions $U \subset S^2$, and they are constructive and provide a representation that attains the optimal approximation rate. Our results also do not rely on a discretization of the transport operator. Unfortunately, the eigenvalue distribution for the spatio-spectral concentration problem on the sphere in Eq. 4.85 remains a conjecture at the moment, and the true boundary function $B(\partial U)$ is currently unknown. However, our experimental results demonstrate the usefulness of the boundary function proposed by us, and we believe it will have applications in other areas where Slepian functions are employed. Additionally, with the equivalence established by us, progress on the problem in other fields will also improve the estimates for the effective dimension of the scattering transport operator $\mathcal{T}$.

We studied the effective dimension of light transport by approximating the local neighborhood $\mathcal{U}' \subset \mathcal{M}$ on the scene manifold by a subset $U \subset S^2$ of the sphere. Other choices for U are conceivable and a subset $U \subset \mathbb{R}^2$ of the plane is sensible in particular for regions $\mathcal{U}' \subset \mathcal{M}$ where curvature is negligible. In the plane, the spectrum of the spatio-spectral concentration problem consists again of the three characteristic parts in Eq. 4.85,[131] and Simons et al.[132] furthermore showed that asymptotically for $A(U) \to 0$ the spatio-spectral concentration problem on the sphere reduces to that in the plane. The consistency of these results is important since any estimate for the effective dimension should be

[131]Slepian, "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty IV: Extensions to many Dimensions; Generalized Prolate Spheroidal Functions".

[132]Simons, Dahlen, and Wieczorek, "Spatiospectral Concentration on a Sphere".

independent of the details of the approximations employed.[133]

In Sec. 4.4.1.2, we derived our result with the axis of radial symmetry being the local normal $\vec{n}(x)$. Our derivation is more generally applicable as long as $\vec{c}(x)$ can be identified uniquely with a point on the sphere. For example for the Phong operator considered by Mahajan et al.,[134] with the reflection direction as the axis of symmetry, the required identification is $(\theta, \phi) = (2\,\theta, \phi)$. Most physically motivated bidirectional reflection distribution functions are not strictly radially symmetric as we assumed in our derivation. The phenomenological success of the Phong model suggests however that radial symmetry is a useful first order approximation, and for the Torrance-Sparrow model, for example, it is known that radial symmetry holds for small outgoing angles.[135].

In our derivation, we assumed that the energy of the L-bandlimited signal is distributed uniformly across bands, that is that the basis function coefficients have the same magnitude independent of $l \leq L$ and completely vanish for $l > L$. Real signals are often better modeled by a soft bandlimit, and for example the small discrepancies between theoretical and empirical spectra in Fig. 4.13 result from a non-uniform energy distribution of the spherical harmonics coefficient for increasing l. To our knowledge the spatio-spectral concentration problem for Sobolev-like spaces with a soft bandlimit has not yet received attention in the literature. In our opinion, this provides an interesting avenue for future work.

The work by Mahajan et al.[136] was an important contribution to the literature. However, we believe that our ansatz, which states the objective as an approximation problem, is more amenable to extensions. For example, spherical harmonics are well suited for the representation of smooth signals. The signals encountered in light transport are however only piece-wise smooth, and hence wavelet-like constructions provide optimal representations.[137] An analysis of the present problem using such bases and nonlinear approximation strategies is another interesting area for future work.

For sampling based techniques, the restricted local variation of light transport has not been exploited systematically in the past. We believe that the understanding gained in this section will be useful to improve the efficacy of these techniques by facilitating the development of rigorous algorithms which exploit coherence between samples.[138] Additionally, applications similar to those considered by Mahajan et al.[139] will benefit from our refined and constructive results.

Slepian basis functions have been studied for Euclidean spaces and on the sphere and Grünbaum and co-workers studied them for general symmetric space.[140] It would be of interest to have an analogue for general Riemannian

[133]We thank Tyler de Witt for discussions on this subject.

[134]Mahajan et al., "A Theory of Locally Low Dimensional Light Transport".

[135]Ramamoorthi and Hanrahan, "A Signal-Processing Framework for Reflection".

[136]Mahajan et al., "A Theory of Locally Low Dimensional Light Transport".

[137]Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*.

[138]The work by Durand and co-workers ("A Frequency Analysis of Light Transport") had a similar objective but we believe that our analysis should provide more concrete guidance.

[139]Mahajan et al., "A Theory of Locally Low Dimensional Light Transport".

[140]Grünbaum, "Eigenvectors of a Toeplitz Matrix: Discrete Version of the Prolate Spheroidal

manifolds, perhaps based on the eigenfunctions of the Laplace-Beltrami operator. For the problem considered in the present section, this would make the assumption that $\mathcal{U}' \subset \mathcal{M}$ can be well approximated by a neighborhood $U \subset S^2$ of the sphere obsolete, and we believe there are many other applications where this would prove useful.

4.4.2   The Functional Analytic Structure of Light Transport

In the previous section, we employed strong assumptions on a light transport scene to obtain a concrete understanding of the outgoing energy density. Desirable are however characterizations that are less restrictive, and which can be verified a priori for a desired application. For example, in Chapter 4.2.3.2 we rationalized the efficiency of our basis projection algorithm by considering the energy distribution of the input signal across spherical harmonics bands. In applications, however, the sought after energy distribution is unavailable, and one would like to predict the performance that can be expected solely based on the information provided by the light transport scene. In fact, such a characterization will be necessary to tailor algorithms towards an application—which will be the key to attain close to optimal performance.

In the following, we will discuss our current understanding of the functional analytic structure of light transport when no assumptions on the scene are made. Although the results are preliminary and hence less conclusive then the foregoing discussions, we believe that the ideas are central for future work, and should hence be presented. We will begin by considering the spectrum of the scattering operator and some of the consequences. In the second part of the section, we will characterize the space of scattered light in each fiber, and we will show that it is a reproducing kernel Hilbert space.

The Spectrum of the Scattering Operator   In a light transport scene $(Q, n, \mathcal{M}, \rho)$, the steady state light energy density $\bar{\ell}$ is given by

$$\bar{\ell} = \ell_0 + \mathcal{T}\ell_0 + \mathcal{T}^2\ell_0 + \dots \tag{4.87}$$

The functional analytic characteristics of $\bar{\ell}$ are hence determined by the properties of the emitted energy density ℓ_0 and those of the scattering transport operator $\mathcal{T}$. We will assume that the properties of ℓ_0 are known and described by a Hilbert space $\mathcal{H}_0$, and we will not consider it any further. Our attention will henceforth be focused on the scattering transport operator $\mathcal{T}$, which will be studied with only few assumptions on $\mathcal{H}_0$, since different spaces will arise in different applications.

The scattering transport operator $\mathcal{T}$ was defined as the concatenation

$$\mathcal{T} = \mathcal{R}_\rho \bar{\mathcal{U}} \, : \, \mathcal{H}(T^+\mathcal{M}) \to \mathcal{H}(T^+\mathcal{M}). \tag{4.88}$$

Wave Functions"; Grünbaum, "Toeplitz Matrices Commuting with Tridiagonal Matrices"; Grünbaum, Longhi, and Perlstadt, "Differential Operators Commuting with Finite Convolution Integral Operators: Some Non-Abelian Examples".

of the surface transport operator $\bar{U} : \mathcal{H}(T^+\mathcal{M}) \to \mathcal{H}(T^-\mathcal{M})$ and the surface scattering operator $\mathcal{R}_\rho : \mathcal{H}(T^-\mathcal{M}) \to \mathcal{H}(T^+\mathcal{M})$. As shown in Proposition 3.2, the surface transport operator is an isometry. In contrast to the light transport operator U_t, however, $\bar{U}$ is no longer unitary, since occlusion breaks the flow of the Hamiltonian vector field, and the rich theory of such operators is hence no longer available. Moreover, the dependence of $\bar{U}$ on the scene geometry makes a generic analysis extremely complex, and we are not able to report progress on the issue. Since it is an integral operator with symmetric kernel, and a Hilbert-Schmidt operator when $\mathcal{H}(T^+\mathcal{M}) \cong \mathcal{H}(T^-\mathcal{M})$ is $L_2(T^+\mathcal{M})$, the scattering operator $\mathcal{R}_\rho$ can be analyzed more easily. By Theorem 2.6, we know that $\mathcal{R}_\rho$ admits a well behaved eigendecomposition with real eigenvalues, and Propositon 3.3 ensures that the largest eigenvalue is strictly bounded by unity.[141] For the moment disregarding the effect of transport and assuming a spatially homogeneous $\mathcal{R}_\rho$, the iteration of the scattering kernel through multiple bounces leads to a characteristic decay in the spectrum of the iterates $\mathcal{R}_\rho^k$ as shown in Fig. 4.14.[142] When the "variation" of the eigenfunctions increases with increasing eigenvalue,[143] the decay corresponds to a smoothing of the light energy density, and through the coercive norm of $\mathcal{R}_\rho$ one obtains a decreasing magnitude. Additionally, the decay also leads to decreasing effective dimension of the scattering operator, that is the "numerical" rank of $\mathcal{R}_\rho$ up to some small $\epsilon > 0$ decreases.[144] To obtain results for the environments of interest, the effect of transport has to be included, in which case occlusion along light transport paths introduces discontinuities. However, the discontinuities are localized, as is well known from images,[145] and the transport operator is hence still "locally unitary", that is along open neighborhoods in phase space the flow remains Hamiltonian. We therefore believe that the above reasoning without transport should locally still be applicable when transport effects are included.

The Space of Scattered Light In the following, we will discuss an alternative characterization of the image space of the scattering operator $\mathcal{R}_\rho$, which will also connect the present considerations again to reproducing kernel Hilbert

---

[141]For general integral operators with symmetric kernel see also Smithies (*Integral Equations*) and Wing (*A Primer on Integral Equations of the First Kind*, Chapter 5), and the overview in (Khvedelidze, *Integral Equation with Symmetric Kernel*).

[142]Stam ("Multiple scattering as a diffusion process") discussed a similar decay already for the scattering events in participating media, and he used it to justify the diffusion approximation for volume transport. The ideas are also central in recent work by Coifman and co-workers ("Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Diffusion Maps"; "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Multiscale Methods"), see in particular the discussion in (Coifman, "Perspectives and Challenges to Harmonic Analysis and Geometry in High Dimensions: Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data").

[143]This behaviour can be described precisely using the Rayleigh quotient and the variational characterization of eigenvalue problems.

[144]Such iterated operators are a standard subject, in particular in the theory of integral equations, see for example (Stakgold and Holst, *Green's Functions and Boundary Value Problems*, Chapter 6).

[145]See for example (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*).
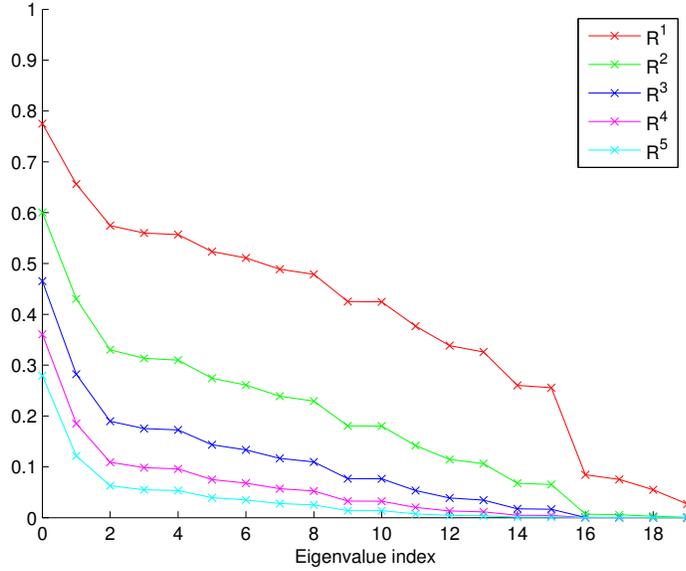
Figure 4.14: Decay of the spectrum of the scattering operator by iteration $R^k = \mathcal{R}_\rho^k$ for a synthetic spectrum with the characteristics of those of the scattering operator.

spaces which played such an important role in the previous part of this chapter. The following result due to Saitoh[146] but adapted for our purposes, enables to tightly characterize the image space of an integral operator.

Theorem 4.1 (Saitoh). *Let X be a set and $L_2(X)$ be the usual Lebesgue space over X, and let $\mathcal{K} : L_2(X) \to \mathcal{H}(X)$ be a linear operator defined by*

$$g = \mathcal{K}\, f \ : \ g(y) = \int_X f(x)\, \kappa(x, y)\, dx$$

with $\kappa(x, y) \in L_2(X \times X)$. Then $(\mathcal{H}, \langle\!\langle\, ,\rangle\!\rangle)$ is a reproducing kernel Hilbert space with kernel

$$k(y, \bar{y}) = \int_X \kappa(x, y)\, \kappa(x, \bar{y})\, dx,$$

and the inner product for $\mathcal{H}$ is $\langle\!\langle f, g \rangle\!\rangle = \langle \bar{F}, \bar{G} \rangle$, where $\bar{F}, \bar{G}$ are the minimum norm pre-images of f, g under $\mathcal{K}$ in $L_2(X)$.

An immediate consequence of the theorem is the following.

Corollary 4.3. *In each fiber, the image space of the scattering operator $\mathcal{R}_\rho$ is a reproducing kernel Hilbert space $(\mathcal{H}_\rho(T_q^+\mathcal{M}), \langle\!\langle\, ,\rangle\!\rangle)$ with kernel*

$$k(\bar{p}, \bar{\bar{p}}) = \int_{T_q^+\mathcal{M}} \rho(p, \bar{p})\, \rho(p, \bar{\bar{p}})\, dp,$$

and the inner product for $\mathcal{H}_\rho$ is as in Theorem 4.1.

---

[146]Saitoh, *Integral Transforms, Reproducing Kernels and their Applications*, Theorem 1.

Note that, by the one-to-one correspondence between a reproducing kernel and its associated Hilbert space, the kernel defined in Corollary 4.3 uniquely identifies $\mathcal{H}_\rho(T_q^+\mathcal{M})$.

Example 4.1. To provide some intuition for Corollary 4.3, we will consider the case of diffuse shading where, up to an inessential constant, the scattering kernel is given by $\rho_d(\bar\omega, \omega) = (\bar\omega \cdot \omega)$, with $\bar\omega$ and ω being considered as unit vectors in $\mathbb{R}^3$. Using Corollary 4.3, we have, by slight abuse of notation in the integral, that

$$k_d(\bar\omega, \tilde\omega) = \int_{S_q^+\mathcal{M}} (\bar\omega \cdot \omega)(\tilde\omega \cdot \omega)\, d\omega. \tag{4.89}$$

Without loss of generality, we can choose $\tilde\omega$ as the normal $n = (0, 0, 1)$. With $\bar\omega = (\bar x, \bar y, \bar z)$ and $\omega = (x, y, z)$ one then obtains

$$k_d(\bar\omega, n) = \int_{S_q^+\mathcal{M}} z(\bar x\, x + \bar y\, y + \bar z\, z)\, d\omega. \tag{4.90}$$

It is easy to verify that only one of the terms is nonzero. Hence, the reproducing kernel for the space of scattered light for a diffuse shading function is given by

$$k_d(\bar\omega, n) = \frac{2\pi}{3}\bar z. \tag{4.91}$$

Unsurprisingly, projecting $k_d(\bar\omega, \tilde\omega)$ into spherical harmonics recovers previous results where diffuse shading was characterized in this domain.[147] In contrast to the reproducing kernel $k_d(\bar\omega, n)$ which uniquely identifies the image space, however, the use of spherical harmonics provides no tight characterization.[148] For example, previously it has been argued that the space of diffuse reflection is essentially nine dimensional.[149] With Slepian functions adapted for the hemisphere, only seven coefficients are need.

Saitoh's theorem, Theorem 4.1, states that the image space $(\mathcal{H}, \langle\!\langle \cdot, \cdot \rangle\!\rangle)$ of a linear integral transform $\mathcal{K}$ is a reproducing kernel Hilbert space whose inner product is

$$\langle\!\langle f, g \rangle\!\rangle = \langle \bar F, \bar G \rangle \tag{4.92}$$

---

[147](Ramamoorthi and Hanrahan, "On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object"; Basri and Jacobs, "Lambertian Reflectance and Linear Subspaces"), and see www.dgp.toronto.edu/people/lessig/dissertation/files/diffuse-sh-projection.nb for the projection of the reproducing kernel $k_d(\bar\omega, n)$ into spherical harmonics.

[148]Spherical harmonics would provide a tight characterization when the scattering kernel would be defined over the sphere $S_q^2\mathcal{M}$ and not only the hemisphere $S_q^+\mathcal{M}$, since any translation invariant operator on S^2 which only depends on $\omega \cdot \bar\omega$ has spherical harmonics as eigenfunctions. A special case of this fact was recently derived in (Mahajan, Tseng, and Ramamoorthi, "An Analysis of the In-Out BRDF Factorization for View-Dependent Relighting"), although the authors there did not understand the structure and generality of their observation.

[149]Ramamoorthi and Hanrahan, "An Efficient Representation for Irradiance Environment Maps".

where $\bar{F}, \bar{G}$ are the minimum norm pre-images of f, g under $\mathcal{K}$, and the inner product on the right hand side is the standard inner product for $L_2(X)$. Saitoh does not provide an explicit expression for the inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$, and it is known that a characterization can be difficult.[150] However, using the eigenvalue decomposition of the scattering kernel

$$\rho(x, y) = \sum_i \lambda_i \, \varphi_i(y) \, \varphi_i(x), \tag{4.93}$$

cf. Remark 2.20, one can show that the inner product for Corollary 4.3 is given by

$$\langle\!\langle f, g \rangle\!\rangle = \left\langle \sum_i \frac{1}{\lambda_i} \, f_i \, \varphi_i(x), \sum_j \frac{1}{\lambda_j} \, g_j \, \varphi_j(x) \right\rangle = \sum_i \frac{1}{\lambda_i^2} \, f_i \, g_i \tag{4.94}$$

which is the standard inner product for $L_2(X)$, with respect to the basis spanned by the φ_i, weighted by $1/\lambda_i^2$. Note the similarity of the inner product with those encountered for Sobolev spaces, cf. Example 2.13. The reproducing property of the kernel in Theorem 4.1 is easily established using Eq. 4.94. By Eq. 4.93, the reproducing kernel can be written as

$$k(y, \bar{y}) = \left\langle \sum_i \sigma_i \, \varphi_i(y) \, \varphi_i(x), \sum_j \sigma_j \, \varphi_j(\bar{y}) \, \varphi_j(x) \right\rangle \tag{4.95a}$$

$$= \sum_{i,j} \sigma_i \, \sigma_j \, \varphi_i(y) \, \varphi_i(\bar{y}) \, \langle \varphi_i(x), \varphi_i(x) \rangle \tag{4.95b}$$

$$= \sum_i \sigma_i^2 \, \varphi_i(y) \, \varphi_i(\bar{y}). \tag{4.95c}$$

and hence we obtain

$$\langle\!\langle f(y), k(y, \bar{y}) \rangle\!\rangle = \left\langle\!\!\left\langle \sum_j f_j \, \varphi_j(y), \sum_i \sigma_i^2 \, \varphi_i(y) \, \varphi_i(\bar{y}) \right\rangle\!\!\right\rangle \tag{4.95d}$$

$$= \sum_i \frac{1}{\sigma_i^2} f_i \left(\sigma_i^2 \, \varphi_i(\bar{y}) \right) \tag{4.95e}$$

$$= \sum_i f_i \, \varphi_i(\bar{y}) \tag{4.95f}$$

$$= f(\bar{y}) \tag{4.95g}$$

which verifies the reproducing property.

[150]See (Saitoh, *Integral Transforms, Reproducing Kernels and their Applications*, Ch. 3) and (Zayed, "A Generalization of the Prolate Spheroidal Wave Functions", Sec. 3).

4.5   Discussion

For light transport simulation in the complex environments encountered in today's applications only a pointwise evaluation of the light energy density is possible. To nonetheless perform effective computations, we introduced reproducing kernel bases, biorthogonal and possibly overcomplete representations whose expansion coefficients are given by function values. These representations eliminate the need to compute inner products, and they provide a practical means to work with continuous functions when only pointwise information is available. We employed reproducing kernel bases to develop finitary point functionals, constructive computational techniques that employ only local information, and together with Galerkin projection they provided an alternative, functional analytic interpretation of algorithms such as distribution ray tracing, path tracing, and photon mapping. This established a common formulation for a wide range of light transport simulation techniques, and it provided insight into the working principles of sampling based techniques with a finite number of samples.

Finitary Point Functionals   Finitary point functionals are computational techniques that employ only local information from a countable set of locations and which admit a formulation using reproducing kernel bases. Many existing techniques for sampling, scattered data approximation, interpolation, and numerical integration are naturally formulated as finitary point functionals, and well known examples which we considered are Monte Carlo integration, Gauss-Legendre quadrature, Lagrange interpolation, and the Shannon sampling theorem.

In contrast to many point functionals in the literature, with a foundation in reproducing kernel bases our finitary point functionals are well defined for finite numerical computations. Moreover, the use of overcomplete representations and the numerical optimization of sampling locations enables to efficiently obtain tight or nearly tight reproducing kernel bases for arbitrary approximation spaces over any domain. This yields finitary point functionals that are close to optimal for the signals in the approximation space,[151] and which markedly outperform classical point functionals such as Monte Carlo and Quasi Monte Carlo techniques that cannot exploit the available structure. The practical benefits are exemplified by numerical integration for which our formulation enables to provide practical answers to the longstanding questions which we already raised in the introduction of the chapter, see the box on the next page.[152]

In many applications, one encounters signals that are not fully contained in the approximation space a finitary point functional was designed for, usually

---

[151]In DeVore's classification, this is optimal for the class of all functions that lie in the chosen approximation space, cf. (DeVore, "Optimal Computation").

[152]See (Dick and Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, p. 16) and (Shirley, Edwards, and Boulos, "Monte Carlo and Quasi-Monte Carlo Methods for Computer Graphics", p. 173).

1. How can we assess the quality of quadrature points?

   The condition number of the reproducing matrix provides a quality measure which is theoretically and practically sound.

2. How can we find quadrature points of particularly high quality?

   Numerical optimization of the sampling points for a kernel basis with the condition number of the reproducing matrix as quality measure provides in many settings a practical means to obtain nearly optimal integration nodes independent of the signal.

3. Should samples on non-square domains be generated on the hypercube or on their native domains?

   Samples should be determined for function spaces natively defined over a problem domain. In fact, the crucial question is for which function space the samples should be designed.

4. Should numeric optimization be used to generate samples?

   Numerical optimization provides the key to obtaining high quality samples, that is samples well adapted to the functional analytic setting of an application.

5. Is there much to be gained from better sampling, or are we already in the diminishing return stage?

   Optimized sampling points, in particular when used for weighted rules, provide the potential to significantly improve the performance attained in applications when these are constructed for a function space that is suitable for the signals of interest.

either because only partial information is available or because the signals are contained in an infinite dimensional function space. We showed that the resulting aliasing error is small when the signal is sufficiently close to the approximation space, and for light transport we demonstrated practical settings where this is satisfied. Additionally, our error analysis provides insight into reproducing locations that minimize the error. This explains the choice of sampling points in classical techniques, and it again enables the use of numerical techniques to find suitable locations for arbitrary domains and function spaces. A key for the robustness of finitary point functionals to aliasing error is the use of overcomplete reproducing kernel bases in which case the well known advantages of frames are exploited.

Many open questions for finitary point functionals were already discussed in Chapter 4.2.4, and the reader is referred there for details. However, some conceptual considerations are in order at this point. We believe that a better understanding of the potential and limitations of finitary point functionals

has to rest on more applications where these are employed. For example, we believe that it is important to better understand the trade-off between a "simple" approximation space, such as the spherical harmonics spaces we employed for radiance probes, and investing more effort in determining the characteristics of a signal and using well adapted representations. Light transport simulation is an interesting point in case in this respect since all information about the problem can in principle be determined. This enables to employ virtually no information, as is currently practice, or to perform an elaborate analysis to find a precise description of all relevant properties, or a middle ground between the two extremes.

Our current examples of finitary points functional were restricted to linear problems, partly because these are the natural choice for a formulation based on linear basis expansions, and partly because computing and working with nonlinearities is much more difficult.[153] Nonetheless, many important nonlinear problems exist, and of relevance for light transport simulation is for example nonlinear approximation, which corresponds to basis projection onto an approximation space that is determined based on the properties of the input signal.[154] It will be interesting to explore how reproducing kernel bases can be useful for the computation of nonlinear problems.

As we discussed before, finitary point functionals are related to a wide range of techniques, for example sampling expansions, multi-grid methods, numerical integration, fast transform algorithms, interpolation schemes, . . . in a multitude of fields, such as approximation theory, information based complexity theory, computational harmonic analysis, applied mathematics, machine learning, computer graphics, . . . Understanding these connections will be well worth the effort to gain more insight into finitary point functionals, and we believe that it will also enable to improve some of the existing techniques in these fields.

A Unified Formulation of Light Transport Simulation The results of Chapter 4.3 demonstrate that reproducing kernel bases enable a functional analytic interpretation of Monte Carlo and density estimation methods, and that this allows for a unified description of a wide range of techniques for light transport simulation based on Galerkin projection, one of the most thoroughly understood and most often used techniques in computational science and engineering. Moreover, our formulation provides insight into the behaviour of techniques such as path tracing and photon mapping with a finite number of samples, and in our opinion this and the known local regularity of the light energy density explain the computational tractability of the techniques observed in practice. Additionally, the formulation presented in Chapter 4.3 rationalizes

---

[153]The work on precomputed radiance transfer that tried to employ nonlinear wavelet approximation is exemplary in this respect, cf. (Ng, Ramamoorthi, and Hanrahan, "All-Frequency Shadows using Non-Linear Wavelet Lighting Approximation"; Ng, Ramamoorthi, and Hanrahan, "Triple Product Wavelet Integrals for All-Frequency Relighting").

[154]See for example (DeVore, "Nonlinear approximation"), and for an introduction (Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Chapter 9).

our belief that not sampling based techniques but our understanding of their working principles was limited in the past. Nonetheless, much more work is necessary to complete the understanding and to gain insight into the function spaces that should be considered, and hence turn our formal exercises into definite statements. Despite these limitations, we believe that the perspective provided by our formulation and the constructive nature of reproducing kernel bases bestow much potential to improve light transport simulation techniques in the future. An interesting challenge will thereby be to break the curse of dimensionality, although we believe that exploiting the local regularity of the energy density and the use of representations that are robust to singularities will allow to do so.

Reproducing kernel Galerkin projection is a general formulation not restricted to light transport simulation, and we believe that it will prove very useful for many other problems in science and engineering where only local information is available.

The Functional Analytic Structure of Light Transport   A central question on the work in this chapter that we have to leave open is the precise functional analytic setting that should be considered: what are the natural function spaces for light transport and which function spaces should be considered for techniques such as distribution ray tracing and path tracing that admit an interpretation using reproducing kernel Galerkin projection?

Our limited current understanding of the functional analytic structure of light transport was discussed in Chapter 4.4. The preliminary results presented there indicate that much of the behaviour of the energy density that has been observed in practice can be explained with the spectrum of the scattering operator. Although the current analysis did not include the effects of transport, we believe that developing the idea of a "locally unitary" operator, which locally along trajectories in phase space preserve the functional analytic structure, will enable to make these arguments precise. In Chapter 4.4, we also showed that reproducing kernel Hilbert spaces arise naturally for light transport as the fiber-wise image space of the scattering operator. More work is however necessary for an interpretation of this result.

Reproducing Kernel Bases and Diffeomorphism Groups   In Proposition 4.2 we established a natural "compatibility" between the pointwise action of a diffeomorphism group on functions and their representation in a reproducing kernel basis. In fact, our technique for the rotation of finite spherical harmonics expansions already provides an implementation of this idea, and it can entirely be interpreted from this perspective. We believe that developing this "compatibility" is a very exciting direction both from a mathematical and a practical point of view, mathematically it might for example provide a different perspective on the functional analytic structure of infinite dimensional Lie groups, and practically it might for instance enable the development of structure preserving integrators for the light energy density, see our discussion in Chapter 3.3.5.

Reproducing Kernel Bases: Why now?   One might ask why reproducing kernel bases have not been considered before in the literature, at least not with the central place they have in our thinking and with the diverse applications for which they are employed by us. Although from today's point of view it is easily overlooked, the rise of biorthogonal bases and frames only began in the 1980s, with the work on wavelets and related representations, and we believe that this provides one answer. Additionally, at least in applied fields, reproducing kernels are largely obscure, and point samples are still commonly interpreted using Dirac deltas.

For us, much of the initial motivation to study the foundations of sampling based techniques came from the diverse applications these have for light transport simulation, and from the difficulties associated with their classical formulation, such as the lacking finite dimensional interpretation of Monte Carlo integration or the inappropriateness of the Shannon sampling theorem for the domains and function space of interest. The origin for our conceptualization of a reproducing kernel basis can be found in a book by Higgins, where orthogonal bases on the sphere formed by reproducing kernel functions are discussed,[155] and it was the rotation of finite spherical harmonics expansions where we developed many of the initial ideas and insights on the use of biorthogonal and overcomplete representations. Vital for our conceptualization was also the interpretation of the characteristic basis as a tight reproducing kernel basis, which enabled to bridge the gap to Monte Carlo integration. Studying the projection of incident light energy density into a basis provided the insight that it can be valuable to employ from the outset a finite dimensional setting but to use overcompletness and numerical optimization to accommodate for real signals that not ideally satisfy our assumptions. It was only after these practical excursions that we developed reproducing kernel Galerkin projection and obtained the unified formulation of algorithms for light transport simulation that was presented in Chapter 4.3.

---

[155](Higgins, *Sampling Theory in Fourier and Signal Analysis: Foundations*), which is based on ideas from (Kempski, *Extension of the Whittaker-Shannon Sampling Series Aided by Symbolic Computation*).

Chapter 5

Conclusion

"Perhaps a lunatic was simply a minority of one."[1]

The present thesis provides a starting point, not an endpoint. We have broken the "arrested development" which persisted for more than 250 years and described a modern theory of light transport. But the theory needs to be filled with life and formulated in its full elegancy. Similarly, for a seemingly disparate set of computational techniques we developed a formulation that is naturally restricted to the pointwise information only available in practice. But the formulation needs to be completed and synthesized into more efficient computational techniques.

From Maxwell's Equations To Light Transport Theory   We began our journey with Maxwell's equations, and we employed microlocal and semi-classical analysis to obtain light transport theory: the description of electromagnetic energy transport on phase space at the short wavelength limit. The electromagnetic energy is then given by the phase space light energy density and time evolution is governed by a canonical Hamiltonian system and described by the light transport equation. Using a Legendre transform, we obtained the Hamiltonian formulation of light transport also from Fermat's principle, and the derivation showed that the time evolution on phase space is a geodesic flow along a natural metric for the system. By considering the homogeneous structure of the Hamiltonian, we established that the dynamics of light transport are equivalently described by a flow on the contact structure of the five dimensional cosphere bundle, fulfilling the prophecy of a tensor description of light transport on a five dimensional phase space.[2] The transport theorem of continuum mechanics enabled us to determine observables for light transport, and this also provided the bridge to classical radiometry, with concept such as radiance and vector irradiance arising from our formulation as quantities to describe measurements.

---

[1]George Orwell, *1984*.

[2]Translators preface of (Gershun, "The Light Field"), see also the beginning of Chapter 3.

> 1. *How can the derivation of light transport theory from Maxwell's equations be formulated geometrically in spacetime?*
>
> 2. *Is ideal light transport an Euler-Arnold system and time evolution described by a geodesic flow on* $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$*?*
>
> 3. *How can light transport be described geometrically when volume and surface scattering is included?*
>
> 4. *What is the intrinsic connection between the action of diffeomorphism groups and reproducing kernel Hilbert spaces?*
>
> 5. *How can tight and nearly tight reproducing kernel bases be characterized?*

In analogy to ideal fluid dynamics, we showed that for a globally defined Hamiltonian vector field ideal light transport is a Lie-Poisson system for the group $\mathrm{Diff}_{\mathrm{can}}(T^*Q)$ of symplectic transformation, and that the light transport equation is then the reduced Hamiltonian equation in the Eulerian representation. It followed that the momentum map for ideal light transport is defined in the convective representation, and computing the map showed that it corresponds to the classical notion of "radiance is constant along a ray", a calculation and conclusion with close resemblance of Kelvin's circulation theorem. We also demonstrated that next to the momentum map additional conserved quantities exist, and these were the analogues of the enstrophy integrals for fluid dynamics. As a Lie-Poisson system, ideal light transport can hence be considered as a phase space analogue of the ideal Euler fluid. By interpreting the Hamiltonian vector field as an anti-self-adjoint operator, Stone's theorem enabled us to describe light transport as a real unitary flow on the space of light energy densities. With this description, we recovered the known operator formulation of light transport by introducing surface transport and scattering operators, and combining these to obtain a description of light propagation in environments with scattering surfaces.

Foundations of Light Transport Simulation Techniques   Our work on the foundations of computational techniques for light transport simulation was motivated by their fundamental restriction to local information: only a pointwise evaluation of the light energy density is possible. To remedy this restriction, we introduced reproducing kernel bases as a computationally and mathematically practical but nonetheless well founded representation to work with pointwise information over arbitrary domains and function spaces. Grounded in these representations, we established finitary point functionals, computational techniques that employ only local information from a countable set of locations. These include for example sampling, pointwise approximation, interpolation, and integration, and we demonstrated that the formulation subsumes many

1. *Can a structure preserving integrator for light transport be obtained by representing the light energy density in a reproducing kernel basis and advecting reproducing points using symplectic integrators?*

2. *Can the transport equation of Poynting's theorem be employed for light transport simulation, and can the angular dependence of the light energy density then be reconstructed using the Wigner transform?*

3. *Can light transport be coupled to other dynamical systems, and does this enable to incrementally update the steady state light energy density?*

4. *What are natural and effective function spaces for the simulation of light transport, and what are bases or frames for these spaces which are enable efficient numerical computations?*

5. *What computational techniques in the literature can be considered as finitary point functionals, and can all practical techniques that employ only local information be formulated as such functionals?*

classical techniques from the literature ranging from Monte Carlo integration over Gauss-Legendre quadrature to Lagrange interpolation. The potential of reproducing kernel bases was exemplified with two applications, the rotation of finite spherical harmonics expansions and the fiber-wise projection of the light energy density into a basis. We demonstrated that close to optimal reproducing kernel bases designed by numerical optimization enable to outperform standard techniques from the literature such as Monte Carlo and Quasi Monte Carlo integration, and the use of overcomplete representations provided the robustness that was needed to ensure that our techniques are effective even when only partial information about the input signal is available or its properties cannot be controlled.

We employed Galerkin projection to derive known techniques for light transport simulation in environments with scattering surfaces. Although the use of Galerkin projection is well known for radiosity, by using reproducing kernel bases we were able to include sampling based techniques such as distribution ray tracing, path tracing, and photon mapping. This provides a unified perspective for a wide range of algorithms in the literature, and it ensures that these admit well defined interpretations with a finite number of samples—the setting unavoidable for practical computations. Unfortunately, our derivations remained formal, and we are currently not able to characterize the function spaces for which these hold.

Towards a Modern Theory of Light Transport Simulation Our work in this thesis provides the foundations for a modern theory of light transport simulation, and it outlines how practical computational techniques provide approximations to Maxwell's equations, see again the diagram on page ii at the beginning of the thesis. However, to achieve our second objective, and to employ scientific understanding for the developments of more effective computational techniques, more effort is needed to establish how it is best used in practice. Ten questions we consider as vital for making progress towards this objective are presented in the boxes in this section, with many more being provided throughout the thesis, and it is our hope that these will receive attention from theoreticians and practitioners alike.

A question where we fell short of our own expectations—and probably those of the reader—and where we presented only very little progress in the foregoing chapters, is visibility:

> *What is a formulation of light transport that allows to efficiently compute visibility?*

We believe that this question deserves special attention: it is arguably the most serious practical impediment for improving the efficiency of computational techniques for light transport simulation, and we believe that a progress on the problem will require alternative mathematical representations that circumvent or alleviate the present difficulties.

Bibliography

Abraham, R. and J. E. Marsden. *Foundations of Mechanics*. second ed. Addison-Wesley Publishing Company, Inc., 1978. URL: http://caltechbook.library.caltech.edu/103/.

Adams, M., T. S. Ratiu, and R. Schmid. *A Lie group structure for Fourier integral operators*. Mar. 1986. URL: http://www.springerlink.com/content/pv72326807k315j1/.

— "A Lie group structure for pseudodifferential operators". In: *Mathematische Annalen* 273.4 (Dec. 1986), pp. 529–551. URL: http://www.springerlink.com/content/g2517k0rp4h6124h/.

Adelson, E. H. and J. R. Bergen. "The Plenoptic Function and the Elements of Early Vision". In: *Computation Models of Visual Processing*. Ed. by M. Landy and J. A. Movshon. MIT Press, 1991, pp. 3–20.

Adler, M. "On a trace functional for formal pseudo-differential operators and the symplectic structure of the Korteweg-deVries type equations". In: *Inventiones Mathematicae* 50.3 (Oct. 1978), pp. 219–248. URL: http://www.springerlink.com/content/n21n3416631457v2/.

Agrachev, A. A. and A. V. Sarychev. "Solid Controllability in Fluid Dynamics". In: *Instability in Models Connected with Fluid Flows I*. New York: Springer, 2007, pp. 1–35. eprint: 0701818v1. URL: http://arxiv.org/abs/math/0701818.

Albeverio, S. and S. Mazzucchi. *Path integral: mathematical aspects*. Jan. 2011. URL: http://www.scholarpedia.org/article/Path_integral:_mathematical_aspects.

Appel, A. "Some Techniques for Shading Machine Renderings of Solids". In: *AFIPS 1968 Spring Joint Comptr. Conf.* 1968, pp. 37–45. URL: http://portal.acm.org/citation.cfm?id=1468082.

— "The Notion of Quantitative Invisibility and the Machine Rendering of Solids". In: *ACM Annual Conference/Annual Meeting* (1967). URL: http://portal.acm.org/citation.cfm?id=806007.

Armentano, D., C. Beltrán, and M. Shub. "Minimizing the discrete logarithmic energy on the sphere: The role of random polynomials". In: *Trans. Amer. Math. Soc.* 363 (2011), pp. 2955–2965. URL: http://www.ams.org/journals/tran/2011-363-06/S0002-9947-2011-05243-8/home.html.

Arnaudon, M., A. B. Cruzeiro, and N. Galamba. "Lagrangian Navier-Stokes flows: a stochastic model". In: *Journal of Physics A: Mathematical and Theoretical* 44.17 (Apr. 2011), p. 175501. URL: http://stacks.iop.org/1751-8121/44/i=17/a=175501.

Arnold, D. N., R. S. Falk, and R. Winther. "Finite element exterior calculus: from Hodge theory to numerical stability". In: *Bulletin (New Series) of the American Mathematical Society* 47 (2010), pp. 281–354. URL: http://www.ams.org/journals/bull/2010-47-02/S0273-0979-10-01278-4/home.html.

Arnold, V. I. "Hamiltonian character of the Euler equations of the dynamics of solids and of an ideal fluid". In: *Uspekhi Mat. Nauk.* 24 (1969), pp. 225–226.

Arnold, V. I. *Lectures on Partial Differential Equations*. Springer, 2004.

— *Mathematical Methods of Classical Mechanics*. second ed. Graduate Texts in Mathematics. Springer, Sept. 1989.

— *On Teaching Mathematics*. Paris, 1997.

— *Ordinary Differential Equations*. Trans. by R. Cooke. third ed. Springer-Textbook. Springer, 1992.

— "Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits". In: *Annales de l'institut Fourier* 16 (1966), pp. 319–361. URL: http://aif.cedram.org/aif-bin/item?id=AIF_1966__16_1_319_0.

— "Sur un principe variationnel pour les découlements stationaires des liquides parfaits et ses applications aux problèmes de stabilité non linéaires". In: *J. Mécanique* 5 (1966), pp. 29–43.

Arnold, V. I., S. M. Gusein-Zade, and A. N. Varchenko. *Singularities of Differentiable Maps: The classification of critical points, caustics and wave fronts*. Vol. 1. Monographs in Mathematics. Birkhäuser, 1985.

Arnold, V. I. and B. A. Khesin. *Topological Methods in Hydrodynamics*. Applied Mathematical Sciences. New York: Springer, 1998.

Aronszajn, N. "Theory of Reproducing Kernels". In: *Transactions of the American Mathematical Society* 68.3 (1950), pp. 337–404. URL: http://www.jstor.org/stable/1990404.

Aronszajn, N. and K. T. Smith. "Characterization of Positive Reproducing Kernels. Applications to Green's Functions." In: *American Journal of Mathematics* 79.3 (1957), pp. 611–622. URL: http://www.jstor.org/stable/2372565.

Arvo, J. "Analytic Methods for Simulated Light Transport". Ph.D. thesis. Yale University, 1995.

— "Applications of Irradiance Tensors to the Simulation of Non-Lambertian Phenomena". In: *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM, 1995, pp. 335–342.

— "The Irradiance Jacobian for Partially Occluded Polyhedral Sources". In: *International Conference on Computer Graphics and Interactive Techniques* (1994). URL: http://portal.acm.org/citation.cfm?id=192250.

— "The Role of Functional Analysis in Global Illumination". In: *Rendering Techniques '95*. Ed. by P. M. Hanrahan and W. Purgathofer. New York: Springer-Verlag, 1995, pp. 115–126.

— "Transfer Equations in Global Illumination". In: *Global Illumination, SIGGRAPH '93 Course Notes*. 1993.

Arvo, J., K. Torrance, and B. Smits. "A Framework for the Analysis of Error in Global Illumination Algorithms". In: *SIGGRAPH '94: Proceedings of the 21st annual conference on computer graphics and interactive techniques*. New York, NY, USA: ACM, 1994, pp. 75–84.

Asimov, I. "The Relativity of Wrong". In: *The Skeptical Inquirer* 14.1 (1989), pp. 35–44.

Atcheson, B., I. Ihrke, W. Heidrich, A. Tevs, D. Bradley, M. Magnor, and H.-P. Seidel. "Time-Resolved 3D Capture of Non-Stationary Gas Flows". In: *ACM Transactions on Graphics* 27.5 (Dec. 2008), p. 1. URL: http://dl.acm.org/citation.cfm?id=1409060.1409085.

Atiyah, M. F., V. K. Patodi, and I. M. Singer. "Spectral asymmetry and Riemannian geometry. III". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 79.01 (Jan. 1976), pp. 71–99. URL: http://journals.cambridge.org/abstract_S0305004100052105.

Bal, G. "Radiative transfer equations with varying refractive index: a mathematical perspective". In: *Journal of the Optical Society of America A* 23.7 (July 2006), p. 1639. URL: http://josaa.osa.org/abstract.cfm?URI=josaa-23-7-1639.

— "Transport Approximations in Partially Diffusive Media". In: *Computational Methods in Transport*. Ed. by F. Graziani. Vol. 48. Lecture Notes in Computational Science and Engineering. Berlin/Heidelberg: Springer-Verlag, 2006, pp. 373–400. URL: http://www.springerlink.com/content/ww2n688410287545.

— "Transport through Diffusive and Nondiffusive Regions, Embedded Objects, and Clear Layers". In: *SIAM Journal on Applied Mathematics* 62.5 (2002), pp. 1677 –1697. URL: http://www.jstor.org/stable/3648731.

Bannai, E. and R. M. Damerell. "Tight spherical designs, I". EN. In: *Journal of the Mathematical Society of Japan* 31.1 (Jan. 1979), pp. 199–207. URL: http://projecteuclid.org/euclid.jmsj/1240319488.

— "Tight Spherical Designs, II". In: *J. London Math. Soc.* s2-21.1 (1980), pp. 13–30.

Barabanenkov, Yu. N., A. G. Vinogradov, Yu. A. Kravtsov, and V. I. Tatarskii. "Application of the theory of multiple scattering of waves to the derivation of the radiation transfer equation for a statistically inhomogeneous medium". In: *Radiophysics and Quantum Electronics (Radiofizika)* 15.12 (Dec. 1972), pp. 1420–1425. URL: http://www.springerlink.com/content/r22r215r8g013l17/.

Barron, A. R. "Universal approximation bounds for superpositions of a sigmoidal function". In: *IEEE Transactions on Information Theory* 39.3 (May 1993), pp. 930–945. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=256500.

Basri, R. and D. W. Jacobs. "Lambertian Reflectance and Linear Subspaces". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.2 (Feb. 2003), pp. 218–233. URL: http://www.computer.org/portal/web/csdl/doi/10.1109/TPAMI.2003.1177153.

Bates, S. and A. Weinstein. *Lectures on the Geometry of Quantization*. Berkeley Mathematics Lecture Notes. American Mathematical Society, 1997. URL: http://math.berkeley.edu/~alanw/GofQ.pdf.

Beckmann, P. and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces*. Norwood, MA: Artech House Inc., 1987.

Belot, G. "The Representation of Time and Change in Mechanics". In: *Philosophy of Physics*. Ed. by J. Butterfield and J. Earman. Handbook of the Philosophy of Science. Amsterdam: North-Holland, 2007, pp. 133–227. URL: http://www.sciencedirect.com/science/article/pii/B9780444515605500051.

Ben-Israel, A. and T. N. E. Greville. *Generalized Inverses*. second ed. CMS books in mathematics. Springer, 2003.

Benamou, J.-D. "Direct computation of multivalued phase space solutions for Hamilton-Jacobi equations". In: *Communications on Pure and Applied Mathematics* 52.11 (Nov. 1999), pp. 1443–1475. URL: http://doi.wiley.com/10.1002/(SICI)1097-0312(199911)52:11<1443::AID-CPA3>3.0.CO;2-Y.

Benedetto, J. J. and P. J. S. G. Ferreira. *Modern Sampling Theory: Mathematics and Applications*. Ed. by J. J. Benedetto and P. J. S. G. Ferreira. Applied and Numerical Harmonic Analysis. Birkhäuser, 2001.

Benedetto, J. J. and M. Fickus. "Finite Normalized Tight Frames". In: *Advances in Computational Mathematics* 18.2 (2003), pp. 357–385. URL: http://www.springerlink.com/content/pxw98648jlt3g863/.

Berger, M., T. Trout, and N. Levit. "Ray tracing mirages". In: *IEEE Computer Graphics and Applications* 10.3 (May 1990), pp. 36–41. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?reload=true&arnumber=55151.

Bergmann, S. "Über die Entwicklung der harmonischen Funktionen der Ebene und des Raumes nach Orthogonalfunktionen". In: *Mathematische Annalen* 86.3-4 (Sept. 1922), pp. 238–271. URL: http://www.springerlink.com/content/j723q65850826128/.

Berlinet, A. and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, 2004.

Besieris, I. M. "Propagation of frequency-modulated pulses in a randomly stratified plasma". In: *Journal of Mathematical Physics* 14.6 (Nov. 1973), p. 704. URL: http://link.aip.org/link/doi/10.1063/1.1666382/html.

Blanco, M. A., M. Flórez, and M. Bermejo. "Evaluation of the rotation matrices in the basis of real spherical harmonics". In: *Journal of Molecular Structure* 419 (1997), pp. 19–27.

Blinn, J. F. "Light Reflection Functions for Simulation of Clouds and Dusty Surfaces". In: *SIGGRAPH Comput. Graph.* 16.3 (1982), pp. 21–29.

— "Models of Light Reflection for Computer Synthesized Pictures". In: *ACM SIGGRAPH Computer Graphics* 11.2 (1977). URL: http://portal.acm.org/citation.cfm?id=563893.

— "What Is a Pixel?" In: *IEEE Computer Graphics and Applications* 25 (2005), pp. 82–87.

Bloch, A., P. S. Krishnaprasad, J. E. Marsden, and T. S. Ratiu. "Dissipation Induced Instabilities". In: *Annales Institute H. Poincaré (Analyse Nonlineaire)* 11.1 (1994), pp. 37–90. URL: http://authors.library.caltech.edu/19797/.

Bodmann, B. G. and P. G. Casazza. "The road to equal-norm Parseval frames". In: *Journal of Functional Analysis* 258.2 (Jan. 2010), pp. 397–420. URL: http://dx.doi.org/10.1016/j.jfa.2009.08.015.

Bolin, M. R. and G. W. Meyer. "A Frequency Based Ray Tracer". In: *SIGGRAPH '95: Proceedings of the 22nd annual conference on computer graphics and interactive techniques*. New York, NY, USA: ACM, 1995, pp. 409–418.

Boltzmann, L. *Vorlesungen über Gastheorie*. Leipzig: J. A. Barth, 1896. URL: http://www.archive.org/details/vorlesungenberg01boltgoog.

Boor, C. de and R. E. Lynch. "On Splines and their Minimum Properties". In: *J. Math. Mech.* 15 (1966), pp. 953–970.

Born, M. and E. Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. seventh ed. Cambridge: Cambridge University Press, 1999.

Bossavit, A. *Computational Electromagnetism: Variational Formulations, Complementarity, Edge Elements*. Academic Press, 1997.

Bossavit, A. and L. Kettunen. "Yee-Like Schemes on Staggered Cellular Grids: a Synthesis Between FIT and FEM Approaches". In: *IEEE Transactions on Magnetics* 36.4 (July 2000), pp. 861–867. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=877580.

Bouguer, P. *Essai d'optique sur la gradation de la lumiere*. Paris: Claude Jombert, 1729.

— *Traité d'Optique sur la gradation de la lumière*. Trans. by W. E. K. Middleton. Paris: H. L. Guerin et L. F. Delatour, 1760.

Bourbaki, N. *Elements of Mathematics: Theory of Sets*. Vol. 1. Springer, 1970.

Brading, K. and E. Castellani. "Symmetries and Invariances in Classical Physics". In: *Philosophy of Physics*. Ed. by J. Butterfield and J. Earman. Handbook of the Philosophy of Science. Amsterdam: North-Holland, 2007, pp. 1331–1367. URL: http://www.sciencedirect.com/science/article/pii/B9780444515605500166.

— "Symmetry and Symmetry Breaking". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2008. 2008. URL: http://plato.stanford.edu/archives/fall2008/entries/symmetry-breaking/.

Brenier, Y. "The Least Action Principle and the Related Concept of Generalized Flows for Incompressible Perfect Fluids". In: *Journal of the American Mathematical Society* 2.2 (1989), pp. 225 –255. URL: http://www.jstor.org/stable/1990977.

Brockett, R. W. "Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems". In: *Proceedings of the 27th IEEE Conference on Decision and Control*. IEEE, 1988, pp. 799–803. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=194420.

— "Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems". In: *Linear Algebra and its Applications* 146.1 (Feb. 1991), pp. 79–91. URL: http://dx.doi.org/10.1016/0024-3795(91)90021-N.

Bungartz, H.-J. and M. Griebel. "Sparse grids". In: *Acta Numerica* 13 (May 2004), pp. 147–269. URL: http://journals.cambridge.org/abstract_S0962492904000182.

Burq, N. "Mesures semi-classique et mesures de défaut". In: *Séminaire Bourbaki* 49.826 (1997), pp. 167–194.

Butterfield, J. "On Symplectic Reduction in Classical Mechanics". In: *The Handbook of Philosophy of Physics*. Ed. by J. Butterfield, J. Earman, D. Gabbay, P. Thagard, and J. Woods. Handbook of the Philosophy of Science. North Holland, 2006, pp. 1–131.

Butzer, P. L., P. J. S. G. Ferreira, J. R. Higgins, S. Saitoh, G. Schmeisser, and R. L. Stens. "Interpolation and Sampling: E.T. Whittaker, K. Ogura and Their Followers". In: *Journal of Fourier Analysis and Applications* (July 2010), pp. 1–35. URL: http://www.springerlink.com/content/k0083h57784u4453/.

Cai, W. and V. Shalaev. *Optical Metamaterials*. New York: Springer, 2010. Chap. 1. URL: http://www.springerlink.com/content/w355x3828k355458/.

Carles, R., C. Fermanian-Kammerer, N. Mauser, and H. Stimming. "On the time evolution of Wigner measures for Schrödinger equations". In: *Communications on Pure and Applied Analysis* 8.2 (Dec. 2008), pp. 559–585. eprint: 0803.0667. URL: http://arxiv.org/abs/0803.0667.

Cartan, H. *Differential Forms*. Dover Publications, 2006.

Cartier, P. and C. DeWitt-Morette. *Functional Integration*. Cambridge Monographs on Mathematical Physics. Cambridge: Cambridge University Press, 2006. URL: http://dx.doi.org/10.1017/CBO9780511535062.

Castor, J. I. *Radiation Hydrodynamics*. Cambridge University Press, 2004.

Castro, A. L. and A. Tromba. *Lecture Notes on Morse Theory*. Tech. rep. University of California, Santa Cruz, 2009.

Cendra, H., J. E. Marsden, and T. S. Ratiu. "Geometric Mechanics, Lagrangian Reduction, and Nonholonomic Systems". In: *Mathematics Unlimited - 2001 and Beyond*. New York: Springer-Verlag, 2001, pp. 221–273. URL: http://resolver.caltech.edu/CaltechAUTHORS:20100727-132349494.

— *Lagrangian Reduction by Stages*. Memoirs of the American Mathematical Society. American Mathematical Society, 2001.

Cendra, H., D. D. Holm, J. E. Marsden, and T. S. Ratiu. "Lagrangian Reduction, the Euler-Poincaré Equations and Semidirect Products". In: *Geometry of Differential Equations*. American Mathematical Society Translations. American Mathematical Society, 1998, pp. 1–25.

Cendra, H., D. D. Holm, M. J. W. Hoyle, and J. E. Marsden. "The Maxwell-Vlasov Equations in Euler-Poincaré Form". In: *Journal of Mathematical Physics* 39.6 (1998), p. 3138. URL: http://jmp.aip.org/jmapaq/v39/i6/p3138_s1.

Cercignani, C. *The Boltzmann Equation and its Applications*. New York: Springer-Verlag, 1988.

Cerezo, E., F. Pérez, X. Pueyo, F. J. Seron, and F. X. Sillion. "A survey on participating media rendering techniques". In: *The Visual Computer* 21.5 (June 2005), pp. 303–328. URL: http://www.springerlink.com/content/l655547336416u82/.

Chandrasekhar, S. *Radiative Transfer*. Dover Publications, June 1960.

Channell, P. J. "Canonical Integration of the Collisionless Boltzmann Equation". In: *Annals of the New York Academy of Sciences* 751.1 (May 1995), pp. 152–161. URL: http://www3.interscience.wiley.com/journal/119241204/abstract.

— *Symplectic integration algorithms*. Intern. Rep. AT-6, ATN-83-9. Los Alamos, New Mexico: Los Alamos National Laboratory, 1983.

— "The beam-beam interaction as a discrete Lie-Poisson dynamical system". In: *AIP Conference Proceedings*. Vol. 214. 1. AIP, Oct. 1990, pp. 464–483. URL: http://link.aip.org/link/?APC/214/464/1.

Channell, P. J. and C. Scovel. "Symplectic integration of Hamiltonian systems". In: *Nonlinearity* 3.2 (1990), p. 231. URL: http://stacks.iop.org/0951-7715/3/i=2/a=001.

Channell, P. J. and J. Scovel. "Integrators for Lie-Poisson dynamical systems". In: *Physica D: Nonlinear Phenomena* 50.1 (May 1991), pp. 80–88. URL: http://dx.doi.org/10.1016/0167-2789(91)90081-J.

Chapman, S. and T. G. Cowling. *The Mathematical Theory of Non-uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases*. Cambridge Mathematical Library. Cambridge University Press, 1960, p. 448.

Chen, M. and J. Arvo. "Theory and Application of Specular Path Perturbation". In: *ACM Transactions on Graphics* 19.4 (2000). URL: http://portal.acm.org/citation.cfm?id=380670.

Cheney, E. W. and W. A. Light. *A Course in Approximation Theory*. AMS Bookstore, 2009.

Cheng, S. *Differentiation Under the Integral Sign with Weak Derivatives*. Tech. rep. 2010. URL: www.gold-saucer.org/math/diff-int/diff-int.pdf.

Chern, S.-S. and C. Chevalley. "Élie Cartan and his Mathematical Work". In: *Bull. Amer. Math. Soc.* 58.2 (1952), pp. 217–250.

Chernoff, P. R. and J. E. Marsden. "On continuity and smoothness of group actions". In: *Bull. Amer. Math. Soc.* 76 (1970), pp. 1044–1049. URL: http://www.ams.org/journals/bull/1970-76-05/S0002-9904-1970-12552-6/.

— *Properties of Infinite Dimensional Hamiltonian Systems*. Vol. 425. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 1974. URL: http://www.springerlink.com/content/978-3-540-07011-5/.

Choquet-Bruhat, Y. and J. E. Marsden. "Solution of the local mass problem in general relativity". In: *Communications in Mathematical Physics (1965-1997)* 51.3 (1976), pp. 283–296. URL: http://projecteuclid.org/euclid.cmp/1103900391.

Chorin, A. J. and J. E. Marsden. *A Mathematical Introduction to Fluid Mechanics*. third ed. Texts in Applied Mathematics. Springer, 1993.

Christensen, O. *An Introduction to Frames and Riesz Bases*. Springer, 2006.

Christensen, P. H. "Adjoints and Importance in Rendering: an Overview". In: *IEEE Transactions on Visualization and Computer Graphics* 9 (2003), pp. 329–340.

— "Hierarchical Techniques for Glossy Global Illumination". Ph.D. thesis. University of Washington, Seattle, Washington, 1995.

Christensen, P. H., D. Lischinski, E. J. Stollnitz, and D. H. Salesin. "Clustering for Glossy Global Illumination". In: *ACM Transactions on Graphics (TOG)* 16.1 (1997). URL: http://portal.acm.org/citation.cfm?id=237748.237749.

Christensen, P. H., E. J. Stollnitz, D. Salesin, and T. DeRose. "Global Illumination of Glossy Environments Using Wavelets and Importance". In: *ACM Trans. Graph.* 15 (1996), pp. 37–71.

Christodoulou, D. "The Euler Equations of Compressible Fluid Flow". In: *Bull. Amer. Math. Soc.* 44 (2007), pp. 581–602. URL: http://www.ams.org/journals/bull/2007-44-04/S0273-0979-07-01181-0/home.html.

Chwolson, O. D. "Grundzüge einer mathematischen Theorie der inneren Diffusion des Lichtes". In: *Bull. l'Acad. Imperiale Sci. St. Pétersbourg* 33 (1889), pp. 221–256. URL: http://books.google.com/books/download/Repertorium_f%C3%BCr_physikalische_technik_f.pdf?id=OODUAAAAMAAJ&output=pdf&sig=ACfU3U1IFEJUa4Tqfs-uva6yHtfHBxFT2g.

Cipriano, F. and A. B. Cruzeiro. "Navier-Stokes Equation and Diffusions on the Group of Homeomorphisms of the Torus". In: *Communications in Mathematical Physics* 275.1 (July 2007), pp. 255–269. URL: http://www.springerlink.com/content/j452n475n105u310/.

Clarke, F. J. J. and D. J. Parry. "Helmholtz Reciprocity: Its Validity and Application to Reflectometry". In: *Lighting Research and Technology* 17 (1985), pp. 1–11.

Clausius, R. "Ueber die Concentration von Wärme- und Lichtstrahlen und die Gränzen ihrer Wirkung". In: *Annalen der Physik und Chemie (Poggendorff's Annalen)* 197.1 (1864), pp. 1–44. URL: http://doi.wiley.com/10.1002/andp.18641970102.

Clebsch, A. "Über eine allgemeine Transformation der hydrodynamischen Gleichungen". de. In: *Journal für die reine und angewandte Mathematik (Crelles Journal)* 54 (Jan. 1857), pp. 293–312.

— "Ueber die Integration der hydrodynamischen Gleichungen." de. In: *Journal für die reine und angewandte Mathematik (Crelles Journal)* 56 (Jan. 1859), pp. 1–10.

Cohen, M. F. and J. Wallace. *Radiosity and Realistic Image Synthesis.* San Diego, CA, USA: Academic Press Professional, Inc., 1993.

Coifman, R. R. "Perspectives and Challenges to Harmonic Analysis and Geometry in High Dimensions: Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data". In: *Perspectives in Analysis.* Vol. 27. Mathematical Physics Studies. Berlin/Heidelberg: Springer-Verlag, 2005, pp. 27–35. URL: http://www.springerlink.com/content/v37ur51467344w45.

Coifman, R. R. and S. Lafon. "Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions". In: *Applied and Computational Harmonic Analysis* 21.1 (July 2006), pp. 31–52. URL: http://dx.doi.org/10.1016/j.acha.2005.07.005.

Coifman, R. R., S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Diffusion Maps". In: *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005), pp. 7426–7431. URL: http://www.pnas.org/content/102/21/7426.abstract.

— "Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Multiscale Methods". In: *Proceedings of the National Academy of Sciences of the United States of America* 102 (2005), pp. 7432–7437. URL: http://www.pnas.org/content/102/21/7432.abstract.

Cook, R. L. "Stochastic Sampling in Computer Graphics". In: *ACM Trans. Graph.* 5 (1986), pp. 51–72.

Cook, R. L., T. Porter, and L. Carpenter. "Distributed Ray Tracing". In: *ACM SIGGRAPH Computer Graphics* 18.3 (1984). URL: http://portal.acm.org/citation.cfm?id=808590.

Cook, R. L. and K. E. Torrance. "A Reflectance Model for Computer Graphics". In: *ACM Transactions on Graphics* 1.1 (Jan. 1982), pp. 7–24. URL: http://portal.acm.org/citation.cfm?id=357290.357293.

Cools, R. and I. H. Sloan. "Minimal cubature formulae of trigonometric degree". In: *Mathematics of Computation* 65 (1996), pp. 1583–1600. URL: http://www.ams.org/journals/mcom/1996-65-216/S0025-5718-96-00767-3/home.html.

Courant, R. and D. Hilbert. *Methoden der Mathematischen Physik*. Die Grundlehren der mathematischen Wissenschaften. Berlin: Julius Springer, 1924. URL: http://resolver.sub.uni-goettingen.de/purl?PPN380672502.

Cruzeiro, A. B. "Hydrodynamics, probability and the geometry of the diffeomorphisms group". In: (Oct. 2008). URL: http://adsabs.harvard.edu/abs/2008arXiv0810.5507B.

Cucker, F. and S. Smale. "On the Mathematical Foundations of Learning". In: *Bull. Amer. Math. Soc.* 39 (2002), pp. 1–49. URL: http://www.ams.org/bull/2002-39-01/S0273-0979-01-00923-5/home.html.

Cucker, F. and D. X. Zhou. *Learning Theory: An Approximation Theory Viewpoint*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2007.

Cui, J. and W. Freeden. "Equidistribution on the Sphere". In: *SIAM J. Sci. Comput.* 18 (1997), pp. 595–609.

Cui, T. J., D. Smith, R. Liu, and D. R. Smith, eds. *Metamaterials: theory, design, and applications*. Springer, 2010. Chap. 1. URL: http://www.springerlink.com/content/tm41n057211x3345/.

Cushman, R. H. and L. M. Bates. *Global Aspects of Classical Integrable Systems*. Birkhäuser Basel, 2004, p. 456. URL: http://www.amazon.com/Global-Aspects-Classical-Integrable-Systems/dp/3764354852.

Cuypers, T., R. Horstmeyer, S. B. Oh, P. Bekaert, and R. Raskar. "Validity of Wigner Distribution Function for ray-based imaging". In: *2011 IEEE International Conference on Computational Photography (ICCP)*. IEEE, Apr. 2011, pp. 1–9. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5753129.

Dachsbacher, C., M. Stamminger, G. Drettakis, and F. Durand. "Implicit Visibility and Antiradiance for Interactive Global Illumination". In: *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)* 26 (Aug. 2007). URL: http://www-sop.inria.fr/reves/Basilic/2007/DSDD07.

Dahmen, W. *Private Communication*. 2011.

Dammertz, S. and A. Keller. "Image Synthesis by Rank-1 Lattices". In: *Monte Carlo and Quasi-Monte Carlo Methods 2006*. Ed. by Alexander Keller, Stefan Heinrich, and Harald Niederreiter. Springer, 2008, pp. 217–236.

Daubechies, I. *Ten Lectures on Wavelets*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992.

Dautray, R and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology*. 2nd print. Vol. 6. Springer, 1993.

de Witt, T. "Fluid Simulation in Bases of Laplacian Eigenfunctions". M.Sc. thesis. University of Toronto, 2011. URL: http://www.dgp.toronto.edu/people/tyler/msc/.

De Witt, T., C. Lessig, and E. Fiume. "Fluid simulation using Laplacian eigenfunctions". In: *ACM Transactions on Graphics* 31.1 (Jan. 2012), pp. 1–11. URL: http://dl.acm.org/citation.cfm?id=2077341.2077351.

Desbrun, M., E. Kanso, and Y. Tong. "Discrete Differential Forms for Computational Modeling". In: *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses*. ACM, 2006, pp. 39–54. URL: http://portal.acm.org/citation.cfm?id=1198666.

Deschamps, G. A. "Electromagnetics and Differential Forms". In: *Proceedings of the IEEE* 69 (1981), pp. 676–696.

DeVore, R. A. "Nonlinear approximation". In: *Acta Numerica* 7 (Jan. 1998), pp. 51–150. URL: http://journals.cambridge.org/abstract_S0962492900002816.

— "Optimal Computation". In: *Proceedings of the International Congress of Mathematicians*. Madrid, Spain, 2006.

DeVore, R. A. and G. G. Lorentz. *Constructive approximation*. Springer, 1993. URL: http://books.google.com/books?id=cDqNW6k7_ZwC&pgis=1.

Dick, J. and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*. Cambridge University Press, 2010.

Dick, J., F. Y. Kuo, F. Pillichshammer, and I. H. Sloan. "Construction algorithms for polynomial lattice rules for multivariate integration". In: *Mathematics of Computation* 74 (2005), pp. 1895–1921. URL: http://www.ams.org/journals/mcom/2005-74-252/S0025-5718-05-01742-4/home.html.

Dippé, M. A. Z. and E. H. Wold. "Antialiasing through Stochastic Sampling". In: *SIGGRAPH '85: Proceedings of the 12th annual conference on computer graphics and interactive techniques*. New York, NY, USA: ACM, 1985, pp. 69–78.

Dolin, L. S. "Beam Description of Weakly Inhomogeneous Wave Fields". In: *Izvestiya Vysshikh Uchebnykh Zzvedenni: Seriya Radiofizika* 7 (1964), pp. 559–563.

Donoho, D. L. "High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality". In: *Mathematical Challenges of the 21st Century*. Los Angeles, CA, USA: American Mathematical Society, 2000. URL: http://www-stat.stanford.edu/~donoho/Lectures/CBMS/Curses.pdf.

Dragulete, O., L. Ornea, and T. S. Ratiu. *Cosphere bundle reduction in contact geometry*. Dec. 2002. URL: http://projecteuclid.org/euclid.jsg/1092749566.

Dragulete, O., T. S. Ratiu, and M. Rodriguez-Olmos. "Singular cosphere bundle reduction". In: *Trans. Amer. Math. Soc.* 359.9 (2007), pp. 4209–4235. URL: http://www.ams.org/journals/tran/2007-359-09/S0002-9947-07-04229-8/home.html.

Driscoll, J. R. and D. M. Healy. "Computing Fourier Transforms and Convolutions on the 2-Sphere". In: *Adv. Appl. Math.* 15 (1994), pp. 202–250.

Duderstadt, J. J. and W. R. Martin. *Transport Theory*. Wiley, New York, 1979.

Duffin, R. J. and A. C. Schaeffer. "A Class of Nonharmonic Fourier Series". In: *Transactions of the American Mathematical Society* 72.2 (Mar. 1952), pp. 341–366. URL: http://www.jstor.org/stable/1990760.

Durand, F., G. Drettakis, and C. Puech. "The 3D Visibility Complex". In: *ACM Transactions on Graphics* 21.2 (2002). URL: http://portal.acm.org/citation.cfm?id=508357.508362.

— "The 3D Visibility Complex: A New Approach to the Problems of Accurate Visibility". In: *Proceedings of the eurographics workshop on Rendering techniques '96* (1996). URL: http://portal.acm.org/citation.cfm?id=275458.275484.

— "The Visibility Skeleton: A Powerful and Efficient Multi-Purpose Global Visibility Tool". In: *International Conference on Computer Graphics and Interactive Techniques* (1997). URL: http://portal.acm.org/citation.cfm?id=258734.258785.

Durand, F., N. Holzschuch, C. Soler, E. Chan, and F. Sillion. "A Frequency Analysis of Light Transport". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2005)* 24 (2005).

Dutré, P., K. Bala, and P. Bekaert. *Advanced Global Illumination*. Natick, MA, USA: A. K. Peters, Ltd., 2006.

Ebin, D. G. "Geodesics on the Symplectomorphism Group". In: *Geom. Funct. Anal.* to appear (2011).

Ebin, D. G. and J. E. Marsden. "Groups of Diffeomorphisms and the Motion of an Incompressible Fluid". In: *The Annals of Mathematics* 92.1 (1970), pp. 102–163. URL: http://www.jstor.org/stable/1970699.

Edelsbrunner, H., M. H. Overmars, and D. Wood. "Graphics in Flatland: A Case Study". In: *Advances in Computing Research* 1 (1983), pp. 35–59.

Egan, K., Y.-T. Tseng, N. Holzschuch, F. Durand, and R. Ramamoorthi. "Frequency Analysis and Sheared Reconstruction for Rendering Motion Blur". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2009)* 28.3 (2009). URL: http://portal.acm.org/citation.cfm?id=1576246.1531399.

Eichhorn, J. "Partial differential equations on closed and open manifolds". In: *Handbook of Global Analysis*. Ed. by D. Krupka and D. Saunders. Amsterdam: Elsevier, 2008, pp. 147–288.

Eichhorn, J. and R. Schmid. "Form preserving diffeomorphisms on open manifolds". In: *Annals of Global Analysis and Geometry* 14.2 (May 1996), pp. 147–176. URL: http://www.springerlink.com/content/u8971237j7p74512/.

Elfouhaily, T. and C.-A. Guérin. "A critical survey of approximate scattering wave theories from random rough surfaces". In: *Waves in Random Media* 14.4 (Oct. 2004), R1–R40. URL: http://www.informaworld.com/10.1088/0959-7174/14/4/R01.

Emmrich, C. and A. Weinstein. "Geometry of the transport equation in multicomponent WKB approximations". In: *Communications in Mathematical Physics* 176.3 (Mar. 1996), pp. 701–711. URL: http://www.springerlink.com/content/p0216232qhx40515.

Enflo, P. "A counterexample to the approximation problem in Banach spaces". In: *Acta Mathematica* 130.1 (July 1973), pp. 309–317. URL: http://www.springerlink.com/content/g124318rp4588310/.

Engel, K.-J. and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Graduate Texts in Mathematics. New York: Springer, 1999.

Etnyre, J. B. "Contact Manifolds". In: *Encyclopedia of Mathematical Physics*. Ed. by J.-P. Francoise, G. L. Naber, and T. S. Tsun. Oxford: Academic Press, 2006, pp. 631–636. URL: http://dx.doi.org/10.1016/B0-12-512666-2/00335-7.

Euclid. *Optica*. Alexandria, around 300.

Evans, L. C. and M. Zworski. *Semiclassical Analysis*. Lecture Notes. Berkeley, CA: University of California, 2011. URL: http://math.berkeley.edu/~zworski/semiclassical.pdf.

Feda, M. "A Monte Carlo Approach for Galerkin Radiosity". In: *The Visual Computer* 12.8 (Oct. 1996), pp. 390–405. URL: http://www.springerlink.com/content/nebuddguead86m99.

Feichtinger, G. H. and T. Werther. "Robustness of Regular Sampling in Sobolev Algebras". In: *Sampling, wavelets, and tomography*. Ed. by J. Benedetto and A. I. Zayed. Birkhäuser, 2004. Chap. 4, pp. 83–114.

Feng, K. "Difference schemes for Hamiltonian formalism and symplectic geometry". In: *Journal of Computational Mathematics* 4.3 (July 1986), pp. 279–289.

Feynman, R. P. "Negative probability". In: *Quantum implications: essays in honour of David Bohm*. Ed. by B. J. Hiley and F. D. Peat. Routledge and Keaton Paul, 1987. Chap. 13, pp. 235–248.

Fickus, M., B. Johnson, K. Kornelson, and K. Okoudjou. "Convolutional frames and the frame potential". In: *Applied and Computational Harmonic Analysis* 19.1 (July 2005), pp. 77–91. URL: http://dx.doi.org/10.1016/j.acha.2005.02.005.

Fischer, A., J. E. Marsden, D. Farnsworth, J. Fink, J. Porter, and A. Thompson. "General relativity as a dynamical system on the manifold $\mathcal{A}$ of Riemannian metrics which cover diffeomorphisms". In: *Methods of Local and Global Differential Geometry in General Relativity*. Ed. by A. Fischer, J. E. Marsden, D. Farnsworth, J. Fink, J. Porter, and A. Thompson. Vol. 14. Lecture Notes in Physics. Berlin, Heidelberg: Springer Berlin Heidelberg, 1972, pp. 176–188. URL: http://www.springerlink.com/content/7765r7537 2002341/.

Fiume, E. *The Mathematical Structure of Raster Graphics*. San Diego: Academic Press, 1989.

Forest, E. and R. Ruth. "Fourth-order symplectic integration". In: *Physica D: Nonlinear Phenomena* 43.1 (May 1990), pp. 105–117.

Frankel, T. *The Geometry of Physics*. Cambridge University Press, 2003.

Freeden, W., T. Gervens, and M. Schreiner. *Constructive Approximation on the Sphere (With Applications to Geomathematics)*. Oxford Sciences Publication. Clarendon Press, Oxford University, 1998.

Freeden, W. and M. Schreiner. *Spherical Functions of Mathematical Geosciences*. Springer, 2009. URL: http://simplelink.library.utoronto.ca/url.cfm/95625.

Friberg, A. T., ed. *Selected Papers on Coherence and Radiometry*. Bellingham, WA, SPIE Optical Engineering Press, 1993.

Gavruta, P. "On some identities and inequalities for frames in Hilbert spaces". In: *Journal of Mathematical Analysis and Applications* 321.1 (Sept. 2006), pp. 469–478. URL: http://dx.doi.org/10.1016/j.jmaa.2005.07.080.

Geiges, H. *An Introduction to Contact Topology*. Cambridge University Press, 2008.

— "Contact Geometry". In: *Handbook of Differential Geometry*. Ed. by F. J. E. Dillen and L. C. A. Verstraelen. Vol. 2. Elsevier, 2006, pp. 315 –382. URL: http://dx.doi.org/10.1016/S1874-5741(06)80008-7.

Gérard, P. "Mesures semi-classiques et ondes de Bloch". In: *Séminaire sur les Équations aux Dérivées Partielles* XVI ().

— "Microlocal defect measures". In: *Communications in Partial Differential Equations* 16.11 (1991), pp. 1761–1794. URL: http://www.tandfonline.com/doi/abs/10.1080/03605309108820822.

Gérard, P., P. A. Markowich, N. J. Mauser, and F. Poupaud. "Homogenization limits and Wigner transforms". In: *Communications on Pure and Applied Mathematics* 50.4 (Apr. 1997), pp. 323–379. URL: http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0312(199704)50:4<323::AID-CPA4>3.0.CO;2-C/abstract.

Gershun, A. "The Light Field". Trans. by P. Moon, G. Timoshenko, and Originally published in Russian (Moscow, 1936). In: *Journal of Mathematics and Physics* 18 (1939), pp. 51–151. URL: http://www.expertunion.ru/knigi/gershun-izbrannyie-trudyi-po-fotometrii-i-svetotehnike.html.

Gibbons, J. "Collisionless Boltzmann Equations and Integrable Moment Equations". In: *Physica D: Nonlinear Phenomena* 3.3 (Aug. 1981), pp. 503–511. URL: http://dx.doi.org/10.1016/0167-2789(81)90036-1.

Gibbons, J., D. D. Holm, and C. Tronci. "Geometry of Vlasov kinetic moments: A bosonic Fock space for the symmetric Schouten bracket". In: *Physics Letters A* 372.23 (June 2008), pp. 4184–4196. URL: http://dx.doi.org/10.1016/j.physleta.2008.03.034.

Gimbutas, Z. and L. Greengard. "A fast and stable method for rotating spherical harmonic expansions". In: *Journal of Computational Physics* 228.16 (2009), pp. 5621–5627. URL: http://linkinghub.elsevier.com/retrieve/pii/S0021999109002691.

Gliklikh, Y. E. "Deterministic viscous hydrodynamics via stochastic processes on groups of diffeomorphisms". In: *Probabilistic Methods in Fluid Dynamics*. Ed. by I. M. Davies. World Scientific, 2003, pp. 179–190.

Goldstein, H., C. P. Poole, and J. L. Safko. *Classical Mechanics*. third ed. Addison Wesley, 2002, p. 638. URL: http://books.google.com/books?id=tJCuQgAACAAJ&pgis=1.

Goldstein, R. A. and R. Nagel. "3-D Visual Simulation". In: *SIMULATION* 16.1 (Jan. 1971), pp. 25–31. URL: http://sim.sagepub.com/cgi/content/abstract/16/1/25.

Golub, G. H. and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in Mathematical Sciences. The Johns Hopkins University Press, 1983.

Golubitsky, M. and I. Stewart. *The Symmetry Perspective: From Equilibrium to Chaos in Phase Space and Physical Space*. Birkhäuser Basel, 2004.

Goodman, D. S. "General Principles of Geometrical Optics". In: *Handbook of Optics*. Ed. by M. Bass. third ed. McGraw-Hill Companies, 2010. Chap. 1, pp. 1.3–1.92.

Goral, C. M., K. E. Torrance, C. P Greenberg, and B. Battaile. "Modeling the Interaction of Light between Diffuse Surfaces". In: *SIGGRAPH Comput. Graph.* 18 (1984), pp. 213–222.

Gordon, W. "Zur Lichtfortpflanzung nach der Relativitätstheorie". In: *Annalen der Physik* 377 (1923), pp. 421–456.

Gortler, S. J., P. Schröder, M. F. Cohen, and P. Hanrahan. "Wavelet Radiosity". In: *Computer Graphics Proceedings, Annual Conference Series, 1993 (ACM SIGGRAPH '93 Proceedings)*. 1993, pp. 221–230.

Gosson, M. A. de. *Symplectic Methods in Harmonic Analysis and in Mathematical Physics*. Pseudo-Differential Operators. Basel: Springer, 2011.

Gotay, M. J. and J. Isenberg. "La Symplectification de la Science". In: *Gazette des Mathématiciens* 54.59–79 (1992). URL: http://www.pims.math.ca/~gotay/Symplectization(E).pdf.

Gräf, M. and D. Potts. "On the computation of spherical designs by a new optimization approach based on fast spherical Fourier transforms". In: *Numerische Mathematik* 119.4 (July 2011), pp. 699–724. URL: http://www.springerlink.com/content/f8h41r6251m21648/.

Gräf, M., D. Potts, and G. Steidl. *Quadrature rules, discrepancies and their relations to halftoning on the torus and the sphere*. Tech. rep. Fakultät für Mathematik, TU Chemnitz, 2011.

Greffet, J.-J. and M. Nieto-Vesperinas. "Field theory for generalized bidirectional reflectivity: derivation of Helmholtz's reciprocity principle and Kirchhoff's law". In: *Journal of the Optical Society of America A* 15.10 (Oct. 1998), p. 2735. URL: http://josaa.osa.org/abstract.cfm?URI=josaa-15-10-2735.

Grinspun, E. "Discrete Elastic Rods and Viscous Threads". In: *Physics/Fields Colloquium 2010-2011*. Toronto, 2010.

Gröchenig, K. "Acceleration of the frame algorithm". In: *IEEE Transactions on Signal Processing* 41.12 (1993), pp. 3331–3340. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=258077.

Groenewold, H. "On the principles of elementary quantum mechanics". In: *Physica* 12.7 (Oct. 1946), pp. 405–460. URL: http://dx.doi.org/10.1016/S0031-8914(46)80059-4.

Gröller, E. "Nonlinear Ray Tracing: Visualizing Strange Worlds". In: *The Visual Computer* 11.5 (May 1995), pp. 263–274. URL: http://www.springerlink.com/content/m06814438 903x103/.

Grünbaum, F. A. "Eigenvectors of a Toeplitz Matrix: Discrete Version of the Prolate Spheroidal Wave Functions". In: *SIAM Journal on Algebraic and Discrete Methods* 2.2 (June 1981), pp. 136–141. URL: http://link.aip.org/link/?SML/2/136/1.

— "Toeplitz Matrices Commuting with Tridiagonal Matrices". In: *J. Linear Alg. and Appl.* 40 (1981), pp. 25–36.

Grünbaum, F. A., L. Longhi, and M. Perlstadt. "Differential Operators Commuting with Finite Convolution Integral Operators: Some Non-Abelian Examples". In: *SIAM Journal on Applied Mathematics* 42.5 (1982), pp. 941 –955. URL: http://www.jstor.org/stable/2101092.

Guillemin, V. and S. Sternberg. *Semi-Classical Analysis*. 2010. URL: http://math.mit.edu/~vwg/semiclassGuilleminSternberg.pdf.

Gümral, H. "Geometry of plasma dynamics. I. Group of canonical diffeomorphisms". In: *Journal of Mathematical Physics* 51.8 (2010), p. 083501. URL: http://link.aip.org/link/JMAPAQ/v51/i8/p083501/s1/html.

Gutierrez, D., F. Seron, A. Munoz, and O. Anson. "Simulation of atmospheric phenomena". In: *Computers & Graphics* 30.6 (Dec. 2006), pp. 994–1010. URL: http://dx.doi.org/10.1016/j.cag.2006.05.002.

Haar, A. "Zur Theorie der Orthogonalen Funktionensysteme". In: *Mathematische Annalen* 69 (1910), pp. 331–371.

Hachisuka, T. and H. W. Jensen. "Stochastic Progressive Photon Mapping". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2009)* 28.5 (Dec. 2009), p. 1. URL: http://portal.acm.org/citation.cfm?id=1661412.1618487.

Hachisuka, T., S. Ogaki, and H. W. Jensen. "Progressive Photon Mapping". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2008)* 27.5 (2008). URL: http://portal.acm.org/citation.cfm?id=1409060.1409083.

Hachisuka, T., W. Jarosz, R. P. Weistroffer, K. Dale, G. Humphreys, M. Zwicker, and H. W. Jensen. "Multidimensional Adaptive Sampling and Reconstruction for Ray Tracing". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2008)* 27.3 (2008), pp. 1–10.

Hairer, E., C. Lubich, and G. Wanner. *Geometric Numerical Integration*. second ed. Springer Series in Computational Mathematics. Springer-Verlag, 2006.

Hamel, G. "Die Lagrange-Eulerschen Gleichungen der Mechanik". In: *Z. Mathematik u. Physik* 50 (1904), pp. 1–57.

— *Theoretische Mechanik*. Springer-Verlag, 1949.

Han, D., M. Z. Nashed, and Q. Sun. "Sampling Expansions in Reproducing Kernel Hilbert and Banach Spaces". In: *Numerical Functional Analysis and Optimization* 30.9-10 (Dec. 2009), pp. 971–987. URL: http://dx.doi.org/10.1080/01630560903408606.

Harris, E. "Radiative Transfer in Dispersive Media". In: *Physical Review* 138.2B (1965), B479–B485. URL: http://prola.aps.org/abstract/PR/v138/i2B/pB479_1.

He, X. D., K. E. Torrance, F. X. Sillion, and D. P. Greenberg. "A Comprehensive Physical Model for Light Reflection". In: *ACM SIGGRAPH Computer Graphics* 25.4 (1991). URL: http://portal.acm.org/citation.cfm?id=122738.

Healy, D. M., D. N. Rockmore, P. J. Kostelec, and S. Moore. "FFTs for the 2-Sphere-Improvements and Variations". In: *Journal of Fourier Analysis and Applications* 9.4 (2003),

498

pp. 341–385. URL: http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1007/s00041-003-0018-9.

Hecht, S., S. Shlaer, and M. H. Pirenne. "Energy, Quanta, and Vision". In: *J. Gen. Physiol.* 25 (1942), pp. 817–819.

Heckbert, P. "Finite Element Methods for Radiosity". In: *ACM SIGGRAPH '93 Course Notes - Global Illumination*. New York, NY, USA: ACM Press, 1993. Chap. 5, pp. 1–7.

— "Introduction to Finite Element Methods". In: *ACM SIGGRAPH '93 Course Notes - Global Illumination*. New York, NY, USA: ACM Press, 1993. Chap. 3, pp. 1–10.

— "Radiosity in Flatland". In: *Computer Graphics Forum* 11.3 (May 1992), pp. 181–192. URL: http://doi.wiley.com/10.1111/1467-8659.1130181.

Heckbert, P. and J. M. Winget. *Finite Element Methods for Global Illumination*. Technical Report UCB/CSD 91/643. Berkeley, CA: University of California at Berkeley, 1991.

Helgason, S. *Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators and Spherical Functions*. 2000.

Helmholtz, H. von. "Die theoretische Grenze für die Leistungsfähigkeit der Mikroskope". In: *Annalen der Physik und Chemie (Poggendorff's Annalen)* 50 (1874), pp. 556–584.

Henon, M. "Vlasov Equation". In: *Astronomy and Astrophysics* 114.1 (1982), pp. 211–212.

Hertz, H. *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*. Ed. by H. v. Helmholtz. Leipzig: Barth, 1894.

Hesse, K. "Complexity of numerical integration over spherical caps in a Sobolev space setting". In: *Journal of Complexity* In press (Sept. 2010). URL: http://dx.doi.org/10.1016/j.jco.2010.09.002.

Hesse, K. and I. H. Sloan. "Cubature over the sphere S^2 x S^2 in Sobolev spaces of arbitrary order". In: *Journal of Approximation Theory* 141.2 (Aug. 2006), pp. 118–133. URL: http://portal.acm.org/citation.cfm?id=1167744.1167746.

— "Worst-case errors in a Sobolev space setting for cubature over the sphere S^2". English. In: *Bulletin of the Australian Mathematical Society* 71.01 (Feb. 2005), pp. 81–105. URL: http://journals.cambridge.org/abstract_S0004972700038041.

Hesse, K., I. H. Sloan, and R. S. Womersley. "Numerical Integration on the Sphere". In: *Handbook of Geomathematics*. Ed. by W. Freeden, Z. M. Nashed, and T. Sonar. Springer-Verlag, 2010.

Hickernell, F. J. "Quadrature Error Bounds with Applications to Lattice Rules". In: *SIAM Journal on Numerical Analysis* 33.5 (1996), pp. 1995–2016. URL: http://www.jstor.org/stable/2158341.

Higgins, J. R. "Sampling Theory for Paley-Wiener Spaces in the Riesz Basis Setting". In: *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences* 94A.2 (1994), pp. 219–236. URL: http://www.jstor.org/stable/20489488.

— *Sampling Theory in Fourier and Signal Analysis: Foundations*. Oxford: Oxford University Press, 1996.

Hilbert, D. "Mathematical Problems". In: *Bulletin of the American Mathematical Society* 8.10 (1902), pp. 437–479.

Holm, D. D. *Geometric Mechanics: Dynamics and Symmetry*. Vol. Part I. London: Imperial College Press, 2008.

— *Geometric mechanics: Rotating, translating and rolling*. Vol. II. Imperial College Press, 2008.

Holm, D. D., W. P. Lysenko, and J. C. Scovel. "Moment invariants for the Vlasov equation". In: *Journal of Mathematical Physics*. Vol. 31. 7. 1990, p. 1610. URL: http://jmp.aip.org/jmapaq/v31/i7/p1610_s1.

Holm, D. D., J. E. Marsden, and T. S. Ratiu. *The Euler-Poincaré Equations and Semidirect Products with Applications to Continuum Theories*. July 1998. URL: http://dx.doi.org/10.1006/aima.1998.1721.

Holm, D. D., V. Putkaradze, and C. Tronci. "Double-bracket dissipation in kinetic theory for particles with anisotropic interactions". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* (Apr. 2010), p. 43. URL: http://rspa.royalsocietypublishing.org/cgi/content/abstract/rspa.2010.0043v1.

— "Geometric dissipation in kinetic equations". In: *Comptes Rendus Mathematique* 345.5 (Sept. 2007), pp. 297–302. URL: http://dx.doi.org/10.1016/j.crma.2007.07.001.

Holm, D. D, V. Putkaradze, and D. Tronci. "Kinetic Models of Heterogeneous Dissipation". In: *Journal of Physics A: Mathematical and Theoretical* 41.34 (Aug. 2008), p. 344010. eprint: 0712.0397. URL: http://arxiv.org/abs/0712.0397.

Holm, D. D., T. Schmah, and C. Stoica. *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*. Oxford texts in applied and engineering mathematics. Oxford University Press, 2009, 515 p.

Holm, D. D. and K. B. Wolf. "Lie-Poisson Description of Hamiltonian Ray Optics". In: *Physica D* 51.1 (1991), p. 189. URL: http://portal.acm.org/citation.cfm?id=120125.

Hörmander, L. *The Analysis of Linear Partial Differential Operators*. Springer, 2004.

Horn, R. A. and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

Howe, M. S. "On the Kinetic Theory of Wave Propagation in Random Media". In: *Philosophical Transactions of the Royal Society of London (Series A)* 274.1242 (1973), pp. 523–549.

Ihrke, I., G. Ziegler, A. Tevs, C. Theobalt, M. Magnor, and H.-P. Seidel. "Eikonal Rendering: Efficient Light Transport in Refractive Objects". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2007)* 26 (2007), p. 59.

Ihrke, I., K. Berger, B. Atcheson, M. Magnor, and W. Heidrich. "Tomographic Reconstruction and Efficient Rendering of Refractive Gas Flows". In: *Imaging Measurement Methods for Flow Analysis*. Ed. by W. Nitsche and C. Dobriloff. Notes on Numerical Fluid Mechanics and Multidisciplinary Design. Berlin Heidelberg: Springer Verlag, 2009, pp. 145 –154.

Ivanic, J. and K. Ruedenberg. "Rotation Matrices for Real Spherical Harmonics. Direct Determination by Recursion". In: *The Journal of Physical Chemistry* 100.15 (Jan. 1996), pp. 6342–6347. URL: http://pubs.acs.org/doi/abs/10.1021/jp953350u.

Jackson, J. D. *Classical Electrodynamics*. third ed. Wiley, Aug. 1998.

James, R. W. "Transformation of Spherical Harmonics Under Change of Reference Frame". In: *Geophysical Journal International* 17.3 (1969), pp. 305–316. URL: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-246X.1969.tb00239.x.

Jensen, H. W. "Importance Driven Path Tracing using the Photon Map". In: *Rendering Techniques*. 1995, pp. 326–335.

— *Realistic Image Synthesis using Photon Mapping*. Natick, MA, USA: A. K. Peters, Ltd., 2001.

Jensen, H. W. and T. Akenine-Möller. "The Race for Real-Time Photorealism". In: *American Scientist* 98.2 (2010), pp. 132–139.

Jensen, H. W. and N. J. Christensen. "Photon Maps in Bidirectional Monte Carlo Ray Tracing of Complex Objects". In: *Computers & Graphics* 19 (1995), pp. 215–224.

Jensen, H. W. and P. H. Christensen. "Efficient simulation of light transport in scences with participating media using photon maps". In: *Proceedings of the 25th annual conference on Computer graphics and interactive techniques - SIGGRAPH '98*. New York, New York, USA: ACM Press, July 1998, pp. 311–320. URL: http://dl.acm.org/citation.cfm?id=280814.280925.

Jost, J. *Postmodern Analysis*. Universitext. Berlin, Heidelberg: Springer-Verlag, 1998.

— *Riemannian Geometry and Geometric Analysis*. Universitext. Springer, 2008.

Kajiya, J. T. "Anisotropic Reflection Models". In: *International Conference on Computer Graphics and Interactive Techniques* 19.3 (1985). URL: http://portal.acm.org/citation.cfm?id=325334.325167.

— "The Rendering Equation". In: *ACM SIGGRAPH Computer Graphics* 20.4 (1986). URL: http://portal.acm.org/citation.cfm?id=15902.

Kajiya, J. T. and B. P. von Herzen. "Ray Tracing Volume Densities". In: *International Conference on Computer Graphics and Interactive Techniques* 18.3 (1984). URL: http://portal.acm.org/citation.cfm?id=800031.808594.

Kambe, T. *Elementary Fluid Mechanics*. World Scientific Publishing, 2007.

Kandrup, H. E. "The secular instability of axisymmetric collisionless star clusters". In: *The Astrophysical Journal* 380 (Oct. 1991), p. 511. URL: http://adsabs.harvard.edu/abs/1991ApJ...380..511K.

Kanso, E., M. Arroyo, Y. Tong, A. Yavari, J. E. Marsden, and M. Desbrun. "On the geometric character of stress in continuum mechanics". In: *Zeitschrift für angewandte Mathematik und Physik* 58.5 (Sept. 2007), pp. 843–856. URL: http://www.springerlink.com/content/8283q278133p8546.

Kautz, J., P.-P. Sloan, and J. Snyder. "Fast, Arbitrary BRDF Shading for Low-Frequency Lighting using Spherical Harmonics". In: *EGSR 2002: Proceedings of the 13th Eurographics Workshop on Rendering*. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2002, pp. 291–296.

Kay, D. S. "Transparency, Refraction, and Ray Tracing for Computer Synthesized Images". M.Sc. thesis. Cornell University, 1979.

Keiner, J. and D. Potts. "Fast evaluation of quadrature formulae on the sphere". In: *Mathematics of Computation* 77 (2008), pp. 397–419.

Keller, A. "Instant Radiosity". In: *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 49–56.

— "Myths of Computer Graphics". In: *Monte Carlo and Quasi-Monte Carlo Methods 2004*. Ed. by Harald Niederreiter and Denis Talay. Springer, 2006, pp. 217–243.

Kempski, B. L. *Extension of the Whittaker-Shannon Sampling Series Aided by Symbolic Computation*. M.Sc. thesis. Anglia Polytechnic University, Cambridge, 1995.

Khesin, B. "Dynamics of symplectic fluids and point vortices". In: (June 2011), p. 12. eprint: 1106.1609. URL: http://arxiv.org/abs/1106.1609.

— "Topological Fluid Dynamics". In: *Notices of the AMS* 52.1 (2005), pp. 9–19.

Khesin, B. and G. Misiolek. "Shock Waves for the Burgers Equation and Curvatures of Diffeomorphism Groups". In: *Proc. Steklov Math. Inst.* 259 (Feb. 2007), pp. 73–81. eprint: 0702196. URL: http://arxiv.org/abs/math/0702196.

Khesin, B., J. Lenells, G. Misiolek, and S. C. Preston. "Geometry of diffeomorphism groups, complete integrability and optimal transport". In: (May 2011). eprint: 1105.0643. URL: http://arxiv.org/abs/1105.0643.

Khesin, B. A. "The Group and Hamiltonian Description of Hydrodynamical Systems". In: *Lectures on Topological Fluid Mechanics*. Ed. by R. L. Ricca. Vol. 1973. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer, 2009, pp. 139–155. URL: http://www.springerlink.com/content/n041746246367131/.

Khesin, B. A. and Yu. V. Chekanov. "Invariants of the Euler equations for ideal or barotropic hydrodynamics and superconductivity in D dimensions". In: *Physica D: Nonlinear Phenomena* 40.1 (Nov. 1989), pp. 119–131. URL: http://dx.doi.org/10.1016/0167-2789(89)90030-4.

Khesin, B. A. and R. Wendt. *The Geometry of Infinite-Dimensional Groups*. Ergebnisse der Mathematik und ihrer Grenzgebiete (Modern Surveys in Mathematics). Berlin; Heidelberg: Springer, 2009.

Khvedelidze, B. V. *Integral Equation with Symmetric Kernel*. 2002. URL: http://www.encyclopediaofmath.org/index.php?title=Integral_equation_with_symmetric_kernel&oldid=15316.

Kidd, R., J. Ardini, and A. Anton. "Evolution of the Modern Photon". In: *American Journal of Physics* 57 (1989), pp. 27–35. URL: http://link.aip.org/link/?AJP/57/27/1.

Kiehn, R. M. and J. F. Pierce. "Intrinsic Transport Theorem". In: *Physics of Fluids* 12.9 (1969), p. 1941. URL: http://pof.aip.org/pfldas/v12/i9/p1941_s1?isAuthorized=no.

Kim, K. and E. Wolf. "Propagation law for Walther's first generalized radiance function and its short-wavelength limit with quasi-homogeneous sources". In: *Journal of the Optical Society of America A* 4.7 (July 1987), p. 1233. URL: http://josaa.osa.org/abstract.cfm?URI=josaa-4-7-1233.

Kim, M.-J. "Verification of the reciprocity theorem". In: *Applied Optics* 27.13 (July 1988), p. 2645. URL: http://ao.osa.org/abstract.cfm?URI=ao-27-13-2645.

Kirchhoff, G. "Ueber den Zusammenhang zwischen Emission und Absorption von Licht und Waerme". In: *Monatsbericht der Akad. d. Wiss. zu Berlin* (1859).

Klein, M. J. *Paul Ehrenfest*. North-Holland Pub. Co., 1970.

Kleinert, H. *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*. fourth ed. World Scientific Publishing Company, 2006.

Kline, M. and I. W. Kay. *Electromagnetic Theory and Geometrical Optics*. John Wiley and Sons, 1965.

Kolar, I., P. W. Michor, and J. Slovak. *Natural Operations in Differential Geometry*. The electronic library of mathematics. Springer-Verlag, 1993. URL: http://www.emis.ams.org/monographs/KSM/.

Kollig, T. and A. Keller. "Efficient Bidirectional Path Tracing by Randomized Quasi-Monte Carlo Integration". In: *Monte Carlo and Quasi-Monte Carlo Methods 2000*. 2002, pp. 245–257.

— "Efficient Bidirectional Path-Tracing by Randomized Quasi-Monte Carlo Integration". In: *Monte Carlo and Quasi-Monte Carlo Methods 2000*. Ed. by K.-T. Fang, F. J. Hickernell, and H. Niederreiter. Berlin: Springer-Verlag, 2002, pp. 290–305.

— "Efficient Multidimensional Sampling". In: *Comput. Graph. Forum* 21 (2002).

Koopman, B. O. "Hamiltonian Systems and Transformations in Hilbert Space". In: *Proceedings of the National Academy of Sciences* 17.5 (1931), pp. 315–318.

Kovacevic, J. and A. Chebira. "Life Beyond Bases: The Advent of Frames (Part I)". In: *Signal Processing Magazine, IEEE* 24.4 (2007), pp. 86 –104. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4286567.

Kovacevic, J., P. L. Dragotti, and V. K. Goyal. "Filter bank frame expansions with erasures". In: *IEEE Transactions on Information Theory* 48.6 (2002), pp. 1439–1450. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1003832.

Kriebel, K.-T. "On the Limited Validity of Reciprocity in Measured BRDFs". In: *Remote Sensing of Environment* 58 (1996), pp. 52–62.

Lafortune, E. P. and Y. D. Willems. "A Theoretical Framework for Physically Based Rendering". In: *Comput. Graph. Forum* 13 (1994), pp. 97–107.

— *Bi-Directional Path Tracing*. Alvor, Portugal, 1993.

— "Rendering participating media with bidirectional path tracing". In: *Rendering Techniques '96 (Proceedings of the Seventh Eurographics Workshop on Rendering)*. Ed. by X. Pueyo and P. Schröder. Dec. 1996, pp. 92–101. URL: http://dl.acm.org/citation.cfm?id=275458.275468.

Lagrange, J. L. *Mécanique analytique*. Paris: Ve Courcier, 1811. URL: http://www.archive.org/details/mcaniqueanalyti05lagrgoog.

Lambert, J. H. and E. Anding. *Photometrie*. Leipzig: Engelmann, 1892. URL: http://www.kuttaka.org/~JHL/L1760a.html.

Lambert, J. H. and D. L. DiLaura. *Photometry or On The Measure and Gradation of Light, Colors, and Shade*. Illuminating Engineering Society of North America, 2001.

Lanczos, C. *Linear Differential Operators*. Dover Publications, 1997.

Landau, H. J. "The Eigenvalue Behavior of Certain Convolution Equations". In: *Trans. Amer. Math. Soc.* 115 (1965), pp. 242–256. URL: http://www.jstor.org/stable/1994267.

Landau, H. J. and H. O. Pollak. "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty III". In: *Bell Syst. Tech. J.* 41 (July 1962), pp. 1295–1336.

— "Prolate Spheroidal Wavefunctions, Fourier Analysis and Uncertainty II". In: *Bell Sys. Tech. J.* 40.1 (1961), pp. 65–84.

Landau, H. J. and H. Widom. "Eigenvalue Distribution of Time and Frequency Limiting". In: *J. of Math. Anal. and App.* 77.2 (1980), pp. 469–481. URL: http://www.sciencedirect.com/science/article/B6WK2-4CRHXX1-3F/2/f9e548bf656ff740071456e596458d28.

Laskar, J. "Large-scale chaos in the solar system". In: *Astronomy and Astrophysics* 287 (1994). URL: http://adsabs.harvard.edu/abs/1994A&A...287L...9L.

Laskar, J. and M. Gastineau. "Existence of collisional trajectories of Mercury, Mars and Venus with the Earth." In: *Nature* 459.7248 (June 2009), pp. 817–9. URL: http://dx.doi.org/10.1038/nature08096.

Lau, C.-W. and K. M. Watson. "Radiation Transport along Curved Ray Paths". In: *Journal of Mathematical Physics* 11.11 (Oct. 1970), p. 3125. URL: http://link.aip.org/link/doi/10.1063/1.1665103/html.

Lax, P. D. *Functional Analysis*. New York: Wiley-Interscience, 2002.

Lehtinen, J. "A Framework for Precomputed and Captured Light Transport". In: *ACM Trans. Graph.* 26 (2007), p. 13.

Lehtinen, J., M. Zwicker, E. Turquin, J. Kontkanen, F. Durand, F. X. Sillion, and T. Aila. "A Meshless Hierarchical Representation for Light Transport". In: *ACM Trans. Graph.* 27 (2008), pp. 1–9.

Leonhardt, U. "Optical Conformal Mapping". In: *Science* 312.5781 (June 2006), pp. 1777–80. URL: http://www.sciencemag.org/cgi/content/abstract/312/5781/1777.

Lessig, C., T. de Witt, and E. Fiume. "Efficient and Accurate Rotation of Finite Spherical Harmonics Expansions". In: *Journal of Computational Physics* 231.2 (Jan. 2012), pp. 243–250. URL: http://dx.doi.org/10.1016/j.jcp.2011.09.014.

Lessig, C. and E. Fiume. "On the Effective Dimension of Light Transport". In: *Computer Graphics Forum* 29.4 (Aug. 2010), pp. 1399–1403. URL: http://doi.wiley.com/10.1111/j.1467-8659.2010.01736.x.

— "SOHO: Orthogonal and Symmetric Haar Wavelets on the Sphere". In: *ACM Transactions on Graphics (TOG)* 27.1 (2008). URL: http://portal.acm.org/citation.cfm?id=1330515.

Levoy, M. "Light Fields and Computational Imaging". In: *Computer* 39.8 (Aug. 2006), pp. 46–55. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1673328.

Levoy, M. and P. Hanrahan. "Light Field Rendering". In: *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM, 1996, pp. 31–42.

Li, S. and W. K. Liu. *Meshfree Particle Methods*. Berlin / Heidelberg: Springer, 2004.

Liboff, R. L. *Kinetic Theory: Classical, Quantum, and Relativistic Descriptions*. New York, NY: Wiley, 1998.

Lie, S. *Geometrie der Berührungstransformationen*. Leipzig: B. G. Teubner, 1896.

— *Theorie der Transformationsgruppen*. Ed. by F. Engel. Vol. 2. Leipzig: B. G. Teubner, 1890.

— "Zur Theorie Partieller Differentialgleichungen". In: *Göttinger Nachrichten* (1872), p. 480.

Liebes, S. "Brightness–On the Ray Invariance of B/n^2". In: *American Journal of Physics* 37.9 (Sept. 1969), p. 932. URL: http://scitation.aip.org/journals/doc/AJPIAS-ft/vol_37/iss_9/932_1.html.

Lions, P.-L. and T. Paul. "Sur les mesures de Wigner". In: *Revista matemática iberoamericana* 9.3 (1993), pp. 553–618. URL: http://dialnet.unirioja.es/servlet/articulo?codigo=2050093&info=resumen.

Littlejohn, R. G. "The semiclassical evolution of wave packets". In: *Physics Reports* 138.4-5 (1986), pp. 193–291. URL: http://dx.doi.org/10.1016/0370-1573(86)90103-1.

Littlejohn, R. G. and W. G. Flynn. "Phase integral theory, coupled wave equations, and mode conversion." In: *Chaos* 2.1 (Jan. 1992), pp. 149–158. URL: http://scitation.aip.org/journals/doc/CHAOEH-ft/vol_2/iss_1/149_1.html.

Liu, X., P.-P. Sloan, H.-Y. Shum, and J. Snyder. "All-Frequency Precomputed Radiance Transfer for Glossy Objects". In: *Eurographics Symposium on Rendering 2004*. 2004.

Lommel, E. "Die Photometrie der diffusen Zurückwerfung". In: *Annalen der Physik* 272.2 (1889), pp. 473–502. URL: http://www3.interscience.wiley.com/journal/112507437/abstract.

Lysenko, W. P. and Z. Parsa. "Beam Matching and Halo Control". In: *17th IEEE Particle Accelerator Conference (PAC 97): Accelerator Science, Technology and Applications*. Vancouver, B.C., Canada, 1997.

Mach, E. *Die Mechanik in ihrer Entwicklung: historisch-kritisch dargestellt*. Internationale Wissenschaftliche Bibliothek. Leipzig: F. A. Brockhaus, 1883. URL: http://http//echo.mpiwg-berlin.mpg.de/ECHOdocuViewfull?url=/mpiwg/online/permanent/einstein_exhibition/sources/Q179XRYG/.

Mahajan, D., Y.-T. Tseng, and R. Ramamoorthi. "An Analysis of the In-Out BRDF Factorization for View-Dependent Relighting". In: *Comput. Graph. Forum (Proceedings of EGSR)* 27 (2008), pp. 1137–1145.

Mahajan, D., I. K. Shlizerman, R. Ramamoorthi, and P. Belhumeur. "A Theory of Locally Low Dimensional Light Transport". In: *ACM Trans. Graph.* 26.3 (Proceedings of ACM SIGGRAPH 2007) (2007), pp. 1–9.

Mallat, S. G. *A Wavelet Tour of Signal Processing*. Second. Academic Press, 1999.

— *A Wavelet Tour of Signal Processing: The Sparse Way*. third ed. Academic Press, 2009.

Mandel, L. and E. Wolf. *Optical Coherence and Quantum Optics*. Cambridge, UK: Cambridge University Press, Sept. 1995.

Marathay, A. S. and J. F. McCalmont. "Diffraction". In: *Handbook of Optics*. Ed. by M. Bass. third ed. McGraw-Hill Companies, 2010. Chap. 3, pp. 3.1–3.38.

Marsden, J. and A. Weinstein. "Reduction of Symplectic Manifolds with Symmetry". In: *Reports on Mathematical Physics* 5.1 (Feb. 1974), pp. 121–130. URL: http://dx.doi.org/10.1016/0034-4877(74)90021-4.

Marsden, J. E. "A Group Theoretic Approach to the Equations of Plasma Physics". In: *Canad. Math. Bull.* 25.2 (1982), pp. 129–142. URL: http://www.cds.caltech.edu/~marsden/bib/1982/06-Ma1982/.

— *Lectures on Mechanics*. Cambridge University Press, 1992. URL: http://authors.library.caltech.edu/21546/1/lom.pdf.

— "Remarks on Geometric Mechanics". In: *Duration and change: fifty years at Oberwolfach*. Springer-Verlag, 1994, pp. 254–274. URL: http://resolver.caltech.edu/CaltechAUTHORS:20100831-075453723.

Marsden, J. E. and T. J. R. Hughes. *Mathematical Foundations of Elasticity*. New York: Dover Publications, 1983. URL: http://resolver.caltech.edu/CaltechBOOK:1983.002.

Marsden, J. E. and P. J. Morrison. "Noncanonical Hamiltonian Field Theory and Reduced MHD". In: *Fluids and plasmas : geometry and dynamics*. Contemporary Mathematics. Providence, R.I: American Mathematical Society, 1984, pp. 133–150.

Marsden, J. E., P. J. Morrison, and A. Weinstein. "The Hamiltonian structure of the BBGKY hierarchy equations". In: *Fluids and plasmas : geometry and dynamics*. Contemporary mathematics. American Mathematical Society, Providence, R.I., 1984, pp. 115–124. URL: http://authors.library.caltech.edu/19767/.

Marsden, J. E. and J. Ostrowski. "Symmetries in Motion: Geometric Foundations of Motion Control". In: *Nonlinear Sci. Today* (1998), pp. 1–21.

Marsden, J. E. and T. S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. third ed. Texts in Applied Mathematics. New York: Springer-Verlag, 2004.

— *Mechanical Systems: Symmetry and Reduction*. 2009.

Marsden, J. E., T. S. Ratiu, and R. Abraham. *Manifolds, Tensor Analysis, and Applications*. third. Applied Mathematical Sciences. New York: Springer-Verlag, 2004.

Marsden, J. E. and J. Scheurle. "Lagrangian reduction and the double spherical pendulum". In: *ZAMP: Zeitschrift für angewandte Mathematik und Physik* 44.1 (Jan. 1993), pp. 17–43. URL: http://www.springerlink.com/content/x726522458q205m7/.

— "The Reduced Euler-Lagrange Equations". In: *Fields Inst. Commun.* 1 (1993), pp. 139–164.

Marsden, J. E. and A. Weinstein. "Coadjoint orbits, vortices, and Clebsch variables for incompressible fluids". In: *Physica D: Nonlinear Phenomena* 7.1-3 (May 1983), pp. 305–323. URL: http://dx.doi.org/10.1016/0167-2789(83)90134-3.

— "The Hamiltonian Structure of the Maxwell-Vlasov equations". In: *Physica D: Nonlinear Phenomena* 4.3 (Mar. 1982), pp. 394–406. URL: http://dx.doi.org/10.1016/0167-2789(82)90043-4.

Marsden, J. E. and M. West. "Discrete Mechanics and Variational Integrators". In: *Acta Numerica* 10 (2001), pp. 357–515.

Marsden, J. E., A. Weinstein, T. S. Ratiu, R. Schmid, and R. G. Spencer. "Hamiltonian Systems with Symmetry, Coadjoint Orbits and Plasma Physics". In: *Atti Acad. Sci. Torino Cl. Sci. Fis. Math. Natur.* 1982, pp. 289–340. URL: http://authors.library.caltech.edu/20083/.

Marsden, J.E. and T. S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems – Online Supplement.* Tech. rep. 1998. URL: http://www.cds.caltech.edu/~marsden/books/Mechanics_and_Symmetry.html.

Martinez, A. *An Introduction to Semiclassical and Microlocal Analysis.* Springer, 2002.

Maslov, V. P. *Théorie des Perturbations et Méthodes Asymptotiques.* Paris: Dunod, 1972.

Maxwell, J. C. "On the Dynamical Theory of Gases". In: *Philosophical Transactions of the Royal Society of London* 157.0 (Jan. 1867), pp. 49–88. URL: http://rstl.royalsocietypublishing.org.

Mazur, S. and S. Ulam. "Sur les transformations isométriques d'espaces vectoriels normés". In: *C. R. Acad. Sci.* 194 (1932), pp. 946–948.

McDuff, D. "Lectures on Groups of Symplectomorphisms". In: (Jan. 2002). eprint: 0201032. URL: http://arxiv.org/abs/math/0201032.

McDuff, D. and D. Salamon. *Introduction to Symplectic Topology.* Oxford University Press, 1998, p. 486.

Mee, C. V. M. van der, M. Z. Nashed, and S. Seatzu. "Sampling Expansions and Interpolation in Unitarily Translation Invariant Reproducing Kernel Hilbert Spaces". In: *Advances in Computational Mathematics* 19.4 (Nov. 2003), pp. 355–372. URL: http://www.springerlink.com/content/w718732418440816/.

Mercer, J. "Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations". In: *Philosophical Transactions of the Royal Society of London (Series A)* 209 (1909), pp. 415–446. URL: http://www.jstor.org/stable/91043.

Meschkowski, H. *Hilbertsche Räume mit Kernfunktion.* Grundlehren der mathematischen Wissenschaften. Berlin: Springer, 1962.

Meyer, K. R. "Symmetries and Integrals in Mechanics". In: *Dynamical Systems.* Ed. by M. M. Peixoto. New York: Academic Press, 1973, pp. 259–272.

Mhaskar, H. N., F. J. Narcowich, and J. D. Ward. "Representing and analyzing scattered data on spheres". In: *Multivariate Approximation and Applications.* Ed. by D. L. D. Leviaton and A. Pinkus. Cambridge University Press, 2001, pp. 44–72.

— "Spherical Marcinkiewicz-Zygmund Inequalities and Positive Quadrature". In: *Math. Comput.* 70 (2001), pp. 1113–1130.

Mihalas, D. and B. Weibel Mihalas. *Foundations of Radiation Hydrodynamics.* New York: Oxford University Press, 1984.

Milne-Thomson, L. M. *Theoretical Aerodynamics.* Dover Publications, 1958.

Mishchenko, M. I. "Gustav Mie and the fundamental concept of electromagnetic scattering by particles: A perspective". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 110.14-16 (Sept. 2009), pp. 1210–1222. URL: http://dx.doi.org/10.1016/j.jqsrt.2009.02.002.

— "Maxwell's equations, radiative transfer, and coherent backscattering: A general perspective". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* 101 (2006), pp. 540–555.

— "Radiative Transfer: A New Look of the Old Theory". In: *Radiative Transfer.* Ed. by M. P. Menguc and N. Selcuk. New York: Begell House, 2007, pp. 1–30.

506

Mishchenko, M. I., L. D. Travis, and A. A. Lacis. *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering*. Cambridge University Press, 2006.

Mishchenko, M. I., V. P. Tishkovets, L. D. Travis, B. Cairns, J. M. Dlugach, L. Liu, V. K. Rosenbush, and N. N. Kiselev. "Electromagnetic scattering by a morphologically complex object: Fundamental concepts and common misconceptions". In: *Journal of Quantitative Spectroscopy and Radiative Transfer* (Apr. 2011). URL: http://dx.doi.org/10.1016/j.jqsrt.2010.03.016.

Misner, C. W., K. S. Thorne, and J. A. Wheeler. *Gravitation*. San Francisco, CA, USA: W. H. Freeman, 1973.

Mitchell, D. P. "Generating Antialiased Images at Low Sampling Densities". In: *SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM, 1987, pp. 65–72.

— "Spectrally Optimal Sampling for Distribution Ray Tracing". In: *SIGGRAPH Comput. Graph.* 25 (1991), pp. 157–164.

Mitchell, D. P. and A. N. Netravali. "Reconstruction filters in computer-graphics". In: *ACM SIGGRAPH Computer Graphics* 22.4 (Aug. 1988), pp. 221–228. URL: http://portal.acm.org/citation.cfm?id=378456.378514.

Moler, C. B. *Numerical Computing with MATLAB*. Society for Industrial Mathematics, 2004.

Montgomery, R. *A Tour of Subriemannian Geometries, Their Geodesics, and Applications*. Mathematical Surveys and Monographs. American Mathematical Society, 2001.

Moon, P. *The Scientific Basis of Illumination Engineering*. New York: Dover Publications, 1961.

Moon, P and D. E. Spencer. *The Photic Field*. Cambridge, MA: MIT Press, 1981.

Moore, E. H. "On Properly Positive Hermitian Matrices". In: *Bull. Amer. Math. Soc.* (1916), pp. 66–67.

— "On the Reciprocal of the General Algebraic Matrix". In: *Bull. Amer. Math. Soc.* 26 (1920), pp. 394–395.

Morrison, P. J. "Hamiltonian description of the ideal fluid". In: *Reviews of Modern Physics* 70.2 (Apr. 1998), pp. 467–521. URL: http://rmp.aps.org/abstract/RMP/v70/i2/p467_1.

— "Thoughts on brackets and dissipation: Old and new". In: *Journal of Physics: Conference Series* 169.1 (May 2009). URL: http://stacks.iop.org/1742-6596/169/i=1/a=012006.

Moyal, A. *Maverick Mathematician: The Life and Science of J.E. Moyal*. Australian National University E Press, 2006. URL: http://epress.anu.edu.au/maverick/html/frames.php.

Moyal, J. E. "Quantum mechanics as a statistical theory". English. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 45.01 (Jan. 1949), pp. 99–124. URL: http://journals.cambridge.org/abstract_S0305004100000487.

Mullen, P., K. Crane, D. Pavlov, Y. Tong, and M. Desbrun. "Energy-Preserving Integrators for Fluid Animation". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2009)* 28.3 (2009), pp. 1–8.

Müller, C. *Spherical Harmonics*. Lecture Notes in Mathematics. Springer, 1966.

Musgrave, F. K. and M. Berger. "A note on ray tracing mirages (comments and author's reply)". In: *IEEE Computer Graphics and Applications* 10.6 (1990), pp. 10–12. URL: http://dl.acm.org/citation.cfm?id=616014.617599.

Mysovskikh, I. P. "On the construction of cubature formulas with fewest nodes". In: *Dokl. Akad. Nauk. SSSR* 178 (1968), pp. 1252–1254.

Nashed, M. Z. and Q. Sun. "Sampling and reconstruction of signals in a reproducing kernel subspace of $L_p(R^d)$". In: *Journal of Functional Analysis* 258.7 (Apr. 2010), pp. 2422–2452. URL: http://dx.doi.org/10.1016/j.jfa.2009.12.012.

Nashed, M. Z. and G. Walter. "General Sampling Theorems for Functions in Reproducing Kernel Hilbert Spaces". In: *Mathematics of Control, Signals, and Systems* 4.4 (1991), pp. 363–390.

Neeb, K.-H. "Towards a Lie theory of locally convex groups". In: *Japanese Journal of Mathematics* 1.2 (Sept. 2006), pp. 291–468. URL: http://www.springerlink.com/content/10128437vr572614/.

Neumann, J. von. "Zur Operatorenmethode In der Klassischen Mechanik". EN. In: *The Annals of Mathematics* 33.3 (Dec. 1932), pp. 587–642. URL: http://www.jstor.org/pss/1968537.

Neumann, L., W. Purgathofer, R. F. Tobler, A. Neumann, P. Elias, M. Feda, and X. Pueyo. "The Stochastic Ray Method for Radiosity". In: *Proceedings of the 6th Eurographics Workshop on Ren- dering*. Dublin: Springer-Verlag, Vienna, 1995, pp. 206–208.

Ng, R., R. Ramamoorthi, and P. Hanrahan. "All-Frequency Shadows using Non-Linear Wavelet Lighting Approximation". In: *ACM Trans. Graph.* 22 (2003), pp. 376–381.

— "Triple Product Wavelet Integrals for All-Frequency Relighting". In: *ACM Trans. Graph.* 23 (2004), pp. 477–487.

Nicodemus, F. E. "Radiance". In: *American Journal of Physics* 31 (1963), pp. 368–377. URL: http://link.aip.org/link/?AJP/31/368/1.

Nicodemus, F. E. and H. J. Kostkowski. *Self Study Manual on Optical Radiation Measurements*. Tech. rep. NBS 910. 1977.

Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1992.

Nishita, T. and E. Nakamae. "Calculation of Interreflections and Its Representation Method". In: *1984 Annual Conference of the Japanese Illumination Engineering Society*. Vol. 63. 1984.

Noether, E. "Invariante Variationsprobleme". In: *Nachr. D. König. Gesellsch. D. Wiss. Zu Göttingen, Math-phys. Klasse* (1918), pp. 235–257.

Noether, E. and M. A. Tavel. *Invariant Variation Problems*. Tech. rep. Mar. 2005. eprint: 0503066. URL: http://arxiv.org/abs/physics/0503066.

Novak, E. and H. Woźniakowski. *Tractability of Multivariate Problems: Linear Information*. Vol. I. MS Tracts in Mathematics. European Mathematical Society Publishing House, 2008.

— *Tractability of Multivariate Problems: Standard Information for Functionals*. Vol. II. EMS Tracts in Mathematics. European Mathematical Society Publishing House, 2010.

Novikov, S. P. and I. A. Taimanov. *Modern Geometric Structures and Fields*. Graduate Studies in Mathematics. American Mathematical Society, 2006.

Oh, S. B., S. Kashyap, R. Garg, S. Chandran, and R. Raskar. "Rendering Wave Effects with Augmented Light Field". In: *Computer Graphics Forum (Proceedings of Eurographics 2010)* 29.2 (June 2010), pp. 507–516. URL: http://doi.wiley.com/10.1111/j.1467-8659.2009.01620.x.

Okayama, H. and I. Ogura. "Experimental verification of nonreciprocal response in light scattering from rough surfaces". In: *Applied Optics* 23.19 (Oct. 1984), p. 3349. URL: http://ao.osa.org/abstract.cfm?URI=ao-23-19-3349.

Omori, H. "On the group of diffeomorphisms on a compact manifold". In: *Proc. Symp. Pure Appl. Math. XV*. Amer. Math. Soc., 1970.

Ortega, J.-P. and T. S. Ratiu. *Momentum Maps and Hamiltonian Reduction*. Progress in Mathematics. Boston, Basel, Berlin: Birkhäuser, 2004, p. 497.

Orti, R., S. Rivière, F. Durand, and C. Puech. "Radiosity for dynamic scenes in flatland with the visibility complex". In: *Computer Graphics Forum* 15.3 (Aug. 1996), pp. 237–248. URL: http://doi.wiley.com/10.1111/1467-8659.1530237.

Otto, F. "The Geometry of Dissipative Evolution Equations: The Porous Medium Equation". In: *Communications in Partial Differential Equations* 26.1 & 2 (Jan. 2001), pp. 101–174. URL: http://www.informaworld.com/10.1081/PDE-100002243.

Outerelo, E. and J. M. Ruiz. *Mapping degree theory*. Graduate studies in mathematics. American Mathematical Soc., 2009.

Ovchinnikov, G. I. and V. I. Tatarskii. "On the problem of the relationship between coherence theory and the radiation-transfer equation". In: *Radiophysics and Quantum Electronics* 15.9 (Sept. 1972), pp. 1087–1089. URL: http://www.springerlink.com/content/v76025p095436404/.

Overbeck, R., C. Donner, and R. Ramamoorthi. "Adaptive Wavelet Rendering". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2009)* 28.5 (2009). URL: http://portal.acm.org/citation.cfm?id=1618452.1618486.

Ovsienko, V. Yu., B. A. Khesin, and Yu. V. Chekanov. "Integrals of the Euler equations of multidimensional hydrodynamics and superconductivity". In: *Journal of Soviet Mathematics* 59.5 (May 1992), pp. 1096–1101. URL: http://www.springerlink.com/content/g7v10r44v2608l23/.

Papanicolaou, G. and L. Ryzhik. "Waves and Transport". In: *Hyperbolic equations and frequency interactions*. Ed. by L. A. Caffarelli and E. Weinan. IAS/Park City mathematics series. Amer. Math. Soc., 1999.

Patterson, E. M. and A. G. Walker. "Riemann Extensions". In: *The Quarterly Journal of Mathematics* 3.1 (1952), pp. 19–28. URL: http://qjmath.oxfordjournals.org/content/3/1/19.full.pdf+html.

Pauly, M., T. Kollig, and A. Keller. "Metropolis Light Transport for Participating Media". In: *Proceedings of the Eurographics Workshop on Rendering Techniques 2000* (2000). URL: http://portal.acm.org/citation.cfm?id=732117.

Pavlov, D., P. Mullen, Y. Tong, E. Kanso, J. E. Marsden, and M. Desbrun. "Structure-preserving discretization of incompressible fluids". In: *Physica D: Nonlinear Phenomena* 240.6 (Mar. 2011), pp. 443–458. URL: http://dx.doi.org/10.1016/j.physd.2010.10.012.

Peacher, J. L. and K. M. Watson. "Doppler Shift in Frequency in the Transport of Electromagnetic Waves through an Underdense Plasma". In: *Journal of Mathematical Physics* 11.4 (Oct. 1970), p. 1496. URL: http://link.aip.org/link/doi/10.1063/1.1665286/html.

Peercy, M. S. "Linear Color Representations for Full Speed Spectral Rendering". In: *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*. New York, NY, USA: ACM, 1993, pp. 191–198.

Pendry, J. B., D. Schurig, and D. R. Smith. "Controlling Electromagnetic Fields". In: *Science* 312.5781 (June 2006), pp. 1780–2. URL: http://www.sciencemag.org/cgi/content/abstract/312/5781/1780.

Penrose, R. "A generalized inverse for matrices". English. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 51.03 (July 1955), pp. 406–413. URL: http://journals.cambridge.org/abstract_S0305004100030401.

Peraiah, A. *An Introduction to Radiative Transfer: Methods and Applications in Astrophysics*. Cambridge: Cambridge University Press, 2002.

Pettini, M. *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*. Interdisciplinary Applied Mathematics. Springer, 2007.

Pharr, M. and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2004.

— *Physically Based Rendering: From Theory to Implementation*. second ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2010.

Phong, B.-T. "Illumination for Computer-Generated Images". Ph.D. thesis. The University of Utah, 1973. URL: http://portal.acm.org/citation.cfm?id=906584.

— "Illumination for Computer Generated Pictures". In: *Communications of the ACM* 18.6 (1975). URL: http://portal.acm.org/citation.cfm?id=360839.

Phong, B.-T. and F. C. Crow. "Improved Rendition of Polygonal Models of Curved Surfaces". In: *Proceedings of the 2nd USA-Japan Computer Conference*. 1975.

Pinchon, D. and P. E. Hoggan. "Rotation Matrices for Real Spherical Harmonics: General Rotations of Atomic Orbitals in Space-Fixed Axes". In: *Journal of Physics A: Mathematical and Theoretical* 40 (2007), pp. 1597–1610. URL: http://stacks.iop.org/1751-8121/40/1597.

Pinkall, U., B. Springborn, and S. Weißmann. "A new doubly discrete analogue of smoke ring flow and the real time simulation of fluid flow". In: *Journal of Physics A: Mathematical and Theoretical* 40.42 (Oct. 2007), pp. 12563–12576. URL: http://stacks.iop.org/1751-8121/40/i=42/a=S04.

Pinkus, A. "Density in Approximation Theory". In: *Surveys in Approximation Theory* 1 (2005), pp. 1–45.

Planck, M. *Theorie der Wärmestrahlung*. second ed. Leipzig: Bahrt, 1913.

Pocchiola, M. "Graphics in Flatland Revisited". In: *Lecture Notes in Computer Science (SWAT 90)*. Ed. by J. Gilbert and R. Karlsson. Vol. 447. Springer Berlin / Heidelberg, 1990, pp. 85–96. URL: http://www.springerlink.com/content/7167805gg871p465/.

Poggio, T. and S. Smale. "The Mathematics of Learning: Dealing with Data". In: *Notices of the American Mathematical Society* 50.5 (2003), pp. 537–544.

Poincaré, H. "Sur une forme nouvelle des équations de la méchanique". In: *C.R. Acad. Sci.* 132 (1901), pp. 369–371.

Polterovich, L. *The Geometry of the Group of Symplectic Diffeomorphisms*. Lectures in Mathematics, ETH Zürich. Basel: Birkhäuser, 2001.

Pomraning, G. C. *The Equations of Radiation Hydrodynamics*. International Series of Monographs in Natural Philosophy. Pergamon Press, 1973.

Preisendorfer, R. W. *Radiative Transfer on Discrete Spaces*. Pergamon Press, 1965.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. New York, NY, USA: Cambridge University Press, 1992.

Rahm, G. de. *Differentiable Manifolds: Forms, Currents, Harmonic Forms*. A Series of Comprehensive Studies in Mathematics. Berlin Heidelberg New York Tokyo: Springer-Verlag, 1984.

Ramamoorthi, R. and P. Hanrahan. "A Signal-Processing Framework for Reflection". In: *ACM Transactions on Graphics (TOG)* 23.4 (2004). URL: http://portal.acm.org/citation.cfm?id=1027416.

— "An Efficient Representation for Irradiance Environment Maps". In: *Proceedings of SIGGRAPH 2001*. SIGGRAPH '01. New York, New York, USA: ACM Press, 2001, pp. 497–500. URL: http://doi.acm.org/10.1145/383259.383317.

— "Frequency Space Environment Map Rendering". In: *Proceedings of ACM SIGGRAPH 2002*. New York, NY, USA: ACM Press, 2002, pp. 517–526.

Ramamoorthi, R. and P. Hanrahan. "On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object". In: *Journal of the Optical Society of America A* 18.10 (Oct. 2001), p. 2448. URL: http://josaa.osa.org/abstract.cfm?URI=josaa-18-10-2448.

Ramamoorthi, R., M. Koudelka, and P. Belhumeur. "A Fourier Theory for Cast Shadows". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.2 (2005). URL: http://portal.acm.org/citation.cfm?id=1038247.

Ramamoorthi, R., D. Mahajan, and P. Belhumeur. "A First-Order Analysis of Lighting, Shading, and Shadows". In: *ACM Trans. Graph.* 26 (2007), p. 2.

Ratiu, T. S. "A Crash Course in Geometric Mechanics". In: *Geometric Mechanics and Symmetry*. Ed. by J. Montaldi. Cambridge: Cambridge University Press, 2005, pp. 23–156. URL: http://ebooks.cambridge.org/chapter.jsf?bid=CBO9780511526367&cid=CBO9780511526367A009.

Ratiu, T. S. and R. Schmid. "The differentiable structure of three remarkable diffeomorphism groups". In: *Mathematische Zeitschrift* 177.1 (Mar. 1981), pp. 81–100. URL: http://www.springerlink.com/content/n07074g213h18511/.

Reimer, M. *Multivariate Polynomial Approximation*. International Series of Numerical Mathematics. Basel: Birkhäuser Verlag, 2003.

Reimer, M. and B. Sündermann. "A Remez-type Algorithm for the Calculation of Extremal Fundamental Systems for Polynomial Spaces on the Sphere". In: *Computing* 37.1 (Mar. 1986), pp. 43–58. URL: http://www.springerlink.com/content/ph70084823325k31/.

Rivera, A. L., S. M. Chumakov, and K. B. Wolf. "Hamiltonian foundation of geometrical anisotropic optics". In: *Journal of the Optical Society of America A* 12.6 (June 1995), pp. 1380–1389. URL: http://josaa.osa.org/abstract.cfm?URI=josaa-12-6-1380.

Rudin, W. *Functional Analysis*. International Series in Pure and Applied Mathematics. New York: McGraw-Hill Inc., 1991.

— *Principles of Mathematical Analysis*. International Series in Pure & Applied Mathematics. McGraw-Hill Publishing Co., Sept. 1976.

Rushmeier, H. E. "Realistic Image Synthesis for Scenes with Radiatively Participating Media". Ph.D. thesis. Ithaca, N.Y.: Cornell University, 1988.

Rushmeier, H. E. and K. E. Torrance. "The Zonal Method for Calculating Light Intensities in the Presence of a Participating Medium". In: *ACM SIGGRAPH Computer Graphics* 21.4 (1987). URL: http://portal.acm.org/citation.cfm?id=37402.37436.

Ruth, R. D. "A Canonical Integration Technique". en. In: *IEEE Transactions on Nuclear Science* 30.4 (Aug. 1983), pp. 2669–2671.

Ryzhik, L., G. Papanicolaou, and J. B. Keller. "Transport equations for elastic and other waves in random media". In: *Wave Motion* 24.4 (1996), p. 43. URL: http://dx.doi.org/10.1016/S0165-2125(96)00021-2.

Saff, E. and A. Kuijlaars. "Distributing many points on a sphere". In: *The Mathematical Intelligencer* 19.1 (1997), pp. 5–11. URL: http://www.springerlink.com/index/10.1007/BF03024331.

Saillard, M. and A. Sentenac. "Rigorous solutions for electromagnetic scattering from rough surfaces". In: *Waves in Random and Complex Media* 11.3 (July 2001), pp. 103–137. URL: http://www.informaworld.com/10.1080/13616670109409779.

Saitoh, S. *Integral Transforms, Reproducing Kernels and their Applications*. Pitman Research Notes in Mathematics. Longman Scientific & Technical, 1997.

— *Theory of Reproducing Kernels and its Applications*. Pitman research notes in mathematics series. Longman Scientific & Technical, 1988.

Sbert, M. "Error and Complexity of Random Walk Monte Carlo Radiosity". In: *IEEE Transactions on Visualization and Computer Graphics* 3.1 (1997). URL: http://portal.acm.org/citation.cfm?id=614360.

— "The Use of Global Random Directions to Compute Radiosity – Global Monte Carlo Techniques". Ph.D. thesis. Universitat Politecnica de Catalunya, Barcelona, Spain, 1996.

Schaback, R. and H. Wendland. "Kernel techniques: From machine learning to meshless methods". In: *Acta Numerica* 15 (May 2006), pp. 543–639. URL: http://journals.cambridge.org/abstract_S0962492906270016.

Schaepman-Strub, G., M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik. "Reflectance Quantities in Optical Remote Sensing –Definitions and Case Studies". In: *Remote Sensing of Environment* 103 (2006), pp. 27–42.

Schmid, R. "Infinite-Dimensional Hamiltonian Systems". In: *Encyclopedia of Mathematical Physics*. Ed. by J.-P. Francoise, G.L. Naber, and T. S. Tsung. Oxford: Academic Press, 2006, pp. 37–44.

— "Infinite Dimensional Lie Groups and Algebras in Mathematical Physics". In: *Advances in Mathematical Physics* (2010).

Schölkopf, B. and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.

Schröder, P., S. J. Gortler, M. F. Cohen, and P. Hanrahan. "Wavelet Projections for Radiosity". In: *Fourth Eurographics Workshop on Rendering*. Paris, France, June 1993, pp. 105–114.

Schuster, A. "Radiation Through a Foggy Atmosphere". In: *Astrophysical Journal* 21 (1905), pp. 1–22.

Schwarzschild, K. *Gesammelte Werke (Collected Works)*. Ed. by H.-H. Voigt. Berlin-Heidelberg-New York: Springer-Verlag, 1992.

— "Über das Gleichgewicht der Sonnenatmosphäre". In: *Nachr. Kgl. Ges. d. Wiss. zu Gött. Math. Phys. Klasse* 195 (1906), pp. 41–53.

Scovel, C. and A. Weinstein. "Finite Dimensional Lie-Poisson Approximations to Vlasov-Poisson Equations". In: *Communications on Pure and Applied Mathematics* 47.5 (May 1994), pp. 683–709. URL: http://www3.interscience.wiley.com/journal/113400041/abstract.

Scully, M. O. and M. Sargent. "The concept of the photon". In: *Physics Today* 25.3 (Mar. 1972), pp. 38–47. URL: http://link.aip.org/link/?PTO/25/38/1.

Serre, D. "Invariants et dégénérescence symplectique de l'équation d'Euler des fluides parfaits incompressibles". In: *C.R. Acad. Sci. (Paris) Ser. A* 298 (1984), p. 349.

Sharpe, R. W. *Differential geometry: Cartan's generalization of Klein's Erlangen program*. Graduate texts in mathematics. Springer, 1997, p. 421.

Shchegrov, A. V., A. A. Maradudin, and E. R. Méndez. "Multiple Scattering of Light from Randomly Rough Surfaces". In: *Progress in Optics* 46 (2004), pp. 117–241. URL: http://dx.doi.org/10.1016/S0079-6638(03)46002-2.

Shirley, P. "Discrepancy as a Quality Measure for Sample Distributions". In: *Proceedings of Eurographics 1991*. 1991.

Shirley, P., D. Edwards, and S. Boulos. "Monte Carlo and Quasi-Monte Carlo Methods for Computer Graphics". In: *Monte Carlo and Quasi-Monte Carlo Methods 2006*. Ed. by A. Keller, S. Heinrich, and H. Niederreiter. Springer, 2008, pp. 167–177.

Shirley, P., B. Wade, P. M. Hubbard, D. Zareski, B. Walter, and D. P. Greenberg. "Global Illumination via Density-Estimation". In: *Proceedings of 6th Workshop on Rendering*. Springer, 1995, pp. 219–230.

512

Siegel, R. and J. R. Howell. *Thermal Radiation Heat Transfer*. fourth ed. Washington, DC, USA: Taylor & Francis, 2001.

Sillion, F. and C. Puech. *Radiosity and Global Illumination*. San Francisco: Morgan Kaufmann Publishers, 1994.

Simo, J. C., J. E. Marsden, and P. S. Krishnaprasad. "The Hamiltonian structure of nonlinear elasticity: The material and convective representations of solids, rods, and plates". In: *Archive for Rational Mechanics and Analysis* 104.2 (1988), pp. 125–183. URL: http://www.springerlink.com/content/j7271524553m3375/.

Simons, F. J. *Personal Communication*. 2010.

— "Slepian Functions and Their Use in Signal Estimation and Spectral Analysis". In: *Handbook of Geomathematics*. Ed. by W. Freeden, Z. M. Nashed, and T. Sonar. 2010.

Simons, F. J., F. A. Dahlen, and M. A. Wieczorek. "Spatiospectral Concentration on a Sphere". In: *SIAM Review* 48.3 (Jan. 2006), pp. 504–536. URL: http://link.aip.org/link/?SIR/48/504/1.

Simons, F. J., J. C. Hawthorne, and C. D. Beggan. "Efficient analysis and representation of geophysical processes using localized spherical basis functions". In: *Wavelets XIII*. Ed. by V. K. Goyal, M. Papadakis, and D. Van De Ville. Vol. 7446. 1. San Diego, CA, USA: SPIE, Aug. 2009, 74460G–15. URL: http://link.aip.org/link/?PSI/7446/74460G/1.

Singer, S. F. *Symmetry in Mechanics: A Gentle, Modern Introduction*. Boston: Birkhäuser, 2001.

Slepian, D. "On Bandwidth". In: *Proceedings of the IEEE* 64.3 (1976). URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1454379.

— "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty IV: Extensions to many Dimensions; Generalized Prolate Spheroidal Functions". In: *Bell Syst. Tech. J.* 43 (1964), pp. 3009–3057.

— "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty V: The Discrete Case". In: *Bell Syst. Tech. J.* 57.2 (1978), pp. 1371–1430.

— "Some Asymptotic Expansions for Prolate Spheroidal Wave Functions". In: *Journal of Mathematics and Physics* 44 (1965), pp. 99–140.

— "Some Comments on Fourier Analysis, Uncertainty and Modeling". In: *SIAM Review* 25.3 (July 1983), pp. 379–393. URL: http://link.aip.org/link/?SIR/25/379/1.

Slepian, D. and H. O. Pollak. "Prolate Spheroidal Wavefunctions Fourier Analysis and Uncertainty I". In: *Bell Sys. Tech. J.* 40.1 (1961), pp. 43–63. URL: http://dx.doi.org/10.2307/2236101.

Sloan, I. H., F. Y. Kuo, and S. Joe. "Constructing Randomly Shifted Lattice Rules in Weighted Sobolev Spaces". In: *SIAM Journal on Numerical Analysis* 40.5 (Nov. 2002), p. 1650. URL: http://link.aip.org/link/?SJNAAM/40/1650/1.

— "On the Step-by-Step Construction of Quasi-Monte Carlo Integration Rules That Achieve Strong Tractability Error Bounds in Weighted Sobolev Spaces". In: *Mathematics of Computation* 71.240 (2002), pp. 1609–1640. URL: http://www.jstor.org/stable/4099966.

Sloan, I. H. and R. Womersley. "Extremal Systems of Points and Numerical Integration on the Sphere". In: *Advances in Computational Mathematics* 21.1 (2004), pp. 107–125. URL: http://www.springerlink.com/content/w2304u80w0542578/.

Sloan, I. H. and H. Woźniakowski. "When Are Quasi-Monte Carlo Algorithms Efficient for High Dimensional Integrals?" In: *Journal of Complexity* 14.1 (Mar. 1998), pp. 1–33.

Sloan, P.-P., J. Kautz, and J. Snyder. "Precomputed Radiance Transfer for Real-Time Rendering in Dynamic, Low-Frequency Lighting Environments". In: *Proceedings of ACM SIGGRAPH 2002*. New York, NY, USA: ACM Press, 2002, pp. 527–536.

Sloan, P.-P., J. Hall, J. Hart, and J. Snyder. "Clustered Principal Components for Precomputed Radiance Transfer". In: *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*. New York, NY, USA: ACM Press, 2003, pp. 382–391.

Smale, S. "Topology and Mechanics I". In: *Inventiones Mathematicae* 10.4 (1970), pp. 305–331. URL: http://www.springerlink.com/content/p5835k07ww862766/.

— "Topology and Mechanics II". In: *Inventiones Mathematicae* 11.1 (1970), pp. 45–64. URL: http://www.springerlink.com/content/gq48683k480qm764/.

Smale, S. and D.-X. Zhou. "Shannon Sampling II: Connections to Learning Theory". In: *Applied and Computational Harmonic Analysis* 19.3 (2005), pp. 285–302. URL: http://linkinghub.elsevier.com/retrieve/pii/S1063520305000540.

Smilie, J. "The Dynamics of Billiard Flows in Rational Polygons". In: *Dynamical Systems, Ergodic Theory, and Applications*. Ed. by Y. G. Sinai. second. Encyclopaedia of Mathematical Sciences. Springer, 2000.

Smith, A. R. *A Pixel Is Not a Little Square*. Tech. rep. Microsoft Corporation, 1995.

Smithies, F. *Integral Equations*. Cambridge Tracts in Mathematics. Cambridge: Cambridge University Press, 1958.

Smolentsev, N. K. "Diffeomorphism groups of compact manifolds". In: *Journal of Mathematical Sciences* 146.6 (Nov. 2007), pp. 6213–6312. URL: http://www.springerlink.com/content/kp663v1n31516059/.

Smolyak, S. A. "Quadrature and interpolation formulas for tensor products of certain classes of functions". In: *Soviet Math. Dokl.* 4 (1963), pp. 240–243.

Snyder, W.C. "Reciprocity of the bidirectional reflectance distribution function (BRDF) in measurements and models of structured surfaces". In: *IEEE Transactions on Geoscience and Remote Sensing* 36.2 (Mar. 1998), pp. 685–691. URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=662750.

Soler, C., K. Subr, F. Durand, N. Holzschuch, and F. Sillion. "Fourier Depth of Field". In: *ACM Transactions on Graphics* 28.2 (2009). URL: http://portal.acm.org/citation.cfm?id=1516522.1516529.

Sommerfeld, A. and J. Runge. "Anwendung der Vektorrechnung auf die Grundlagen der geometrischen Optik". In: *Annalen der Physik* 340.7 (1911), pp. 277–298. URL: http://doi.wiley.com/10.1002/andp.19113400705.

Song, G. and H. Zhang. "Reproducing Kernel Banach Spaces with the ℓ_1 Norm II: Error Analysis for Regularized Least Square Regression." In: *Neural Computation* 23.10 (Oct. 2011), pp. 2713–29. URL: http://www.mitpressjournals.org/doi/abs/10.1162/NECO_a_00178?journalCode=neco.

Song, G., H. Zhang, and F. J. Hickernell. "Reproducing Kernel Banach Spaces with the ℓ_1 Norm". In: (Jan. 2011). eprint: 1101.4388. URL: http://arxiv.org/abs/1101.4388.

Spanier, J. and E. M. Gelbard. *Monte Carlo Principles and Neutron Transport Problems*. Reading, MA: Addison-Wesley, 1969.

Sparber, C., P. A. Markowich, and N. J. Mauser. "Wigner functions versus WKB-methods in multivalued geometrical optics". In: *Asymptotic Analysis* 33.2 (2003), pp. 153–187.

Spivak, M. *A Comprehensive Introduction to Differential Geometry*. third ed. Vol. 1. Publish or Perish, 1999.

Stakgold, I. and M. J. Holst. *Green's Functions and Boundary Value Problems*. third ed. Pure and Applied Mathematics. John Wiley & Sons, Inc., 2011.

Stam, J. "Multiple scattering as a diffusion process". In: *Proceedings of the 6th Eurographics Workshop on Rendering*. Ed. by P. M. Hanrahan and W. Purgathofer. Dublin, Ireland: Springer Verlag, 1995, pp. 51–58.

Stam, J. and E. Languénou. "Ray Tracing in Non-Constant Media". In: *Proceedings of the Eurographics Workshop on Rendering Techniques '96*. London, UK: Springer-Verlag, 1996, p. 225.

Starck, J.-L., Y. Moudden, P. Abrial, and M. Nguyen. "Wavelets, Ridgelets and Curvelets on the Sphere". en. In: *Astronomy and Astrophysics* 446.3 (Feb. 2006), pp. 1191–1204. URL: http://adsabs.harvard.edu/abs/2006A&A...446.1191S.

Stern, A., Y. Tong, M. Desbrun, and J. E. Marsden. "Geometric Computational Electrodynamics with Variational Integrators and Discrete Differential Forms". In: (2007). eprint: 0707.4470v3. URL: http://arxiv.org/abs/0707.4470.

Stern, D. "Classification Of Magnetic Shells". In: *J. Geophys. Res.* 70.15 (1965), pp. 3629–3634. URL: http://www.agu.org/journals/ABS/1965/JZ070i015p03629.shtml.

Stillwell, J. *Naive Lie Theory*. Undergraduate Texts in Mathematics. Springer, 2008.

Stix, T. H. *The Theory of Plasma Waves*. McGraw-Hill Advanced Physics Monograph Series. McGraw-Hill, 1962, p. 283.

Stone, A. H. "Paracompactness and product spaces". In: *Bulletin (New Series) of the American Mathematical Society* 54.10 (Oct. 1948), pp. 977–982. URL: http://projecteuclid.org/euclid.bams/1183512390.

Stott, P. E. "A transport equation for the multiple scattering of electromagnetic waves by a turbulent plasma". In: *Journal of Physics A: General Physics* 1.6 (Nov. 1968), pp. 675–689. URL: http://stacks.iop.org/0022-3689/1/i=6/a=306.

Sudarshan, E. C. G. and N. Mukunda. *Classical Dynamics: A Modern Perspective*. John Wiley & Sons, 1974.

Synge, J. L. "Focal Properties of Optical and Electromagnetic Systems". In: *The American Mathematical Monthly* 51.4 (1944), pp. 185–200. URL: http://www.jstor.org/stable/2305772.

— "On the Geometry of Dynamics". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 226 (1927), pp. 31–106. URL: http://www.jstor.org/stable/91165.

Szegö, G. *Orthogonal Polynomials*. American Mathematical Society, 1939, p. 432.

Tabellion, E. and A. Lamorlette. "An approximate global illumination system for computer generated films". In: *ACM Transactions on Graphics* 23.3 (Aug. 2004), p. 469. URL: http://portal.acm.org/citation.cfm?id=1015706.1015748.

Tao, T. *Structure and Randomness: Pages from Year One of a Mathematical Blog*. American Mathematical Society, 2008.

— *The Euler-Arnold equation*. 2010. URL: http://terrytao.wordpress.com/2010/06/07/the-euler-arnold-equation/.

Tartar, L. "H-measures, a New Approach for Studying Homogenization, Oscillations and Concentration Effects in Partial Differential Equations". In: *Proc. Roy. Soc. Edin.* 115A (1990), pp. 193–230.

— "H-measures and Applications". In: *International Congress of Mathematicians*. Kyoto: Springer, 1990, pp. 1215–1223.

— "H-measures and Small Amplitude Homogenization". In: *Random Media and Composites*. Ed. by R. V. Kohn and G. W. Milton. Philadelphia, PA, USA: SIAM, 1989, pp. 89–99.

— "Mathematical Tools for Studying Oscillations and Concentrations: from Young Measures to H-Measures and Their Variants". In: *Multiscale problems in science and technology: challenges to mathematical analysis and perspectives*. Ed. by N. Antonić. Springer, 2002. Chap. 1, pp. 1–84.

— *The General Theory of Homogenization: a personalized introduction*. Lecture notes of the Unione Matematica Italiana. Springer, 2010.

Thirring, W. *A Course in Mathematical Physics 1 and 2*. New York Wien: Springer-Verlag, 1992.

Thomas, G. E. and K. Stamnes. *Radiative Transfer in the Atmosphere and Ocean*. Cambridge University Press, 2002, p. 517.

Torrance, K. E. and E. M. Sparrow. "Theory for Off-Specular Reflection From Roughened Surfaces". In: *Journal of the Optical Society of America* 57.9 (Sept. 1967), p. 1105. URL: http://www.opticsinfobase.org/abstract.cfm?URI=josa-57-9-1105.

Traub, J. F. and A. G. Werschulz. *Complexity and Information*. New York, NY, USA: Cambridge University Press, 1999.

Tremaine, S. "Is the solar system stable?" In: *Frontiers of Astrophysics, Proceedings of the Rosseland Centenary Symposium*. Ed. by P. B. Lilje and P. Maltby. Oslo: Institute of Theoretical Astrophysics, 1995, p. 85.

Triebel, H. *Theory of Function Spaces*. reprint. Modern Birkhäuser Classics. Springer, 2010.

Tronci, C. "Geometric dynamics of Vlasov kinetic theory and its moments". Ph.D. Imperial College London, 2008.

Trowbridge, T. S. and K. P. Reitz. "Average Irregularity Representation of a Rough Surface for Ray Reflection". In: *Journal of the Optical Society of America* 65.5 (May 1975), p. 531. URL: http://www.opticsinfobase.org/abstract.cfm?URI=josa-65-5-531.

Tsang, L. and J. A. Kong. *Scattering of Electromagnetic Waves*. Ed. by J. A. Kong. Wiley Interscience, 2000.

Tucker, R. W. "Differential Form Valued Forms and Distributional Electromagnetic Sources". In: *ArXiv e-prints* (2008).

Turski, L. and A. N. Kaufman. "Canonical-dissipative formulation of relativistic plasma kinetic theory with self-consistent Maxwell field". In: *Physics Letters A* 120.7 (Mar. 1987), pp. 331–333. URL: http://dx.doi.org/10.1016/0375-9601(87)90725-0.

Unser, M. "Sampling—50 Years After Shannon". In: *Proceedings of the IEEE* 88 (Apr. 2000), pp. 569–587.

Vale, R. and S. Waldron. "Tight Frames and Their Symmetries". In: *Constructive Approximation* 21.1 (July 2004), pp. 83–112. URL: http://www.springerlink.com/content/4jlk4dxuj4tlp0h5/.

Vallis, G. K., G. F. Carnevale, and W. R. Young. "Extremal energy properties and construction of stable solutions of the Euler equations". In: *J. Fluid Mech.* 207 (1989), pp. 133–152. URL: http://adsabs.harvard.edu//abs/1989JFM...207..133V.

van Hove, L. *On Certain Unitary Representations of an Infinite Group of Transformations*. Trans. by M. Berg and C. DeWitt-Morette. Sur certaines représentations unitaires d'un groupe infini de transformations (Proc. Roy. Acad. Sci. Belgium (26), 1âĂŞ102). Singapore ; River Edge, NJ: World Scientific, 1951.

— "Sur le probléme des relations entre les transformations unitaires de la Mécanique quantique et les transformations canoniques de la Mécanique classique." In: *Acad. Roy. Belg. Bull. Classe Sci. Mem.* 37.5 (1951), pp. 610–620.

Veach, E. "Robust Monte Carlo Methods for Light Transport Simulation". Ph.D. thesis. Stanford University, 1997.

516

Veach, E. and L. J. Guibas. "Bidirectional Estimators for Light Transport". In: *Fifth Euro-graphics Workshop on Rendering*. 1994, pp. 147–162.

— "Metropolis Light Transport". In: *SIGGRAPH '97: Proceedings of the 24th annual conference on computer graphics and interactive techniques*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 65–76.

Venable, W. H. "Comments on reciprocity failure". In: *Applied Optics* 24.23 (Dec. 1985), p. 3943. URL: http://ao.osa.org/abstract.cfm?URI=ao-24-23-3943.

Villani, C. "A Review of Mathematical Topics in Collisional Kinetic Theory". In: *Handbook of Mathematical Fluid Dynamics*. Ed. by S. Friedlander and D. Serre. Elsevier, 2002.

Viskanta, R. *Radiative Transfer of Combustion Systems: Fundamentals and Applications*. Begell House, 2005, p. 454.

Vogelaére, R. de. *Methods of integration which preserve the contact transformation property of the Hamiltonian equations*. Tech. rep. Notre Dame, Indiana: Department of Mathematics, University of Notre Dame, 1956.

Wahba, G. *An Introduction to Model Building With Reproducing Kernel Hilbert Spaces*. course notes. University of Wisconsin-Madison, 2000. URL: http://www.stat.wisc.edu/~wahba/ftp1/interf/index.html.

— *Spline Models for Observational Data*. CBMS-NSF regional conference series in applied mathematics. SIAM, 1990, p. 169.

Walter, B., P. M. Hubbard, P. Shirley, and D. P. Greenberg. "Global Illumination using Local Linear Density Estimation". In: *ACM Trans. Graph.* 16 (1997), pp. 217–259.

Walter, B., S. Fernandez, A. Arbree, K. Bala, M. Donikian, and D. P. Greenberg. "Lightcuts: A Scalable Approach to Illumination". In: *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*. New York, NY, USA: ACM Press, 2005, pp. 1098–1107.

Walter, B., A. Arbree, K. Bala, and D. P. Greenberg. "Multidimensional Lightcuts". In: *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*. New York, NY, USA: ACM Press, 2006, pp. 1081–1088.

Walter, G. G. "Wavelets and Sampling". In: *Modern Sampling Theory: Mathematics and Applications*. Ed. by J. J. Benedetto and P. J. S. G. Ferreira. Applied and Numerical Harmonic Analysis. Birkhäuser, 2001, pp. 49–72.

Walther, A. "Radiometry and Coherence". In: *J. Opt. Soc. Am.* 58 (1968), pp. 1256–1259. URL: http://www.opticsinfobase.org/abstract.cfm?URI=josa-58-9-1256.

Ward, G. J. "The RADIANCE lighting simulation and rendering system". In: *ACM SIG-GRAPH Computer Graphics*. New York, New York, USA: ACM Press, July 1994, pp. 459–472. URL: http://portal.acm.org/citation.cfm?id=192161.192286.

Ward, G. J. and P. Heckbert. "Irradiance Gradients". In: *1992 Eurographics Workshop on Rendering*. Bristol, 1992, pp. 85–98.

Ward, G. J., F. M. Rubinstein, and R. D. Clear. "A ray tracing solution for diffuse interreflec-tion". In: *ACM SIGGRAPH Computer Graphics*. Vol. 22. 4. Aug. 1988, pp. 85–92. URL: http://portal.acm.org/citation.cfm?id=378456.378490.

Warnick, K. F. and D. V. Arnold. "Electromagnetic Green Functions using Differential Forms". In: *Journal of Electromagnetic Waves and Applications* 10 (1996), pp. 427–438.

Warnock, J. E. *A Hidden Line Algorithm for Halftone Picture Representation*. Tech. rep. Salt Lake City, Utah, USA: Utah University, 1968.

Watanabe, Y. "Differential geometry on diffeomorphism groups and Lagrangian stability of viscous flows". In: *Physica D: Nonlinear Phenomena* 225.2 (Jan. 2007), pp. 197–203. URL: http://dx.doi.org/10.1016/j.physd.2006.10.011.

Watson, K. M. "Electromagnetic Wave Scattering within a Plasma in the Transport Approximation". In: *Physics of Fluids* 13.10 (Aug. 1970), p. 2514. URL: http://link.aip.org/link/doi/10.1063/1.1692821/html.

— "Multiple Scattering of Electromagnetic Waves in an Underdense Plasma". In: *Journal of Mathematical Physics* 10.4 (Nov. 1969), p. 688. URL: http://link.aip.org/link/doi/10.1063/1.1664895/html.

Weinberg, S. "Eikonal Method in Magnetohydrodynamics". In: *Physical Review* 126.6 (1962), pp. 1899–1909. URL: http://prola.aps.org.myaccess.library.utoronto.ca/abstract/PR/v126/i6/p1899_1.

Weißmann, S. and U. Pinkall. "Filament-based smoke with vortex shedding and variational reconnection". In: *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2010)* 29.4 (2010). URL: http://portal.acm.org/citation.cfm?id=1833349.1778852.

Weyl, H. "Quantenmechanik und Gruppentheorie". In: *Zeitschrift für Physik* 46.1-2 (Nov. 1927), pp. 1–46. URL: http://www.springerlink.com/content/j0246228q3j8v74g/.

— "Über die Gleichverteilung von Zahlen mod. Eins". In: *Mathematische Annalen* 77.3 (Sept. 1916), pp. 313–352. URL: http://www.springerlink.com/content/g7843g50wx7ln388/.

Whitted, T. "An Improved Illumination Model for Shaded Display". In: *Communications of the ACM* 23 (1980), pp. 343–349.

Wiener, N. "Generalized Harmonic Analysis". In: *Acta Mathematica* 55.1 (Dec. 1930), pp. 117–258. URL: http://www.springerlink.com/content/h604825553g4j183/.

Wigner, E. *Gruppentheorie und ihre Anwendungen auf die Quantenmechanik der Atomspektren.* Braunchweig: Vieweg Verlag, 1931.

— "On the Quantum Correction For Thermodynamic Equilibrium". In: *Phys. Rev.* 40.5 (June 1932), pp. 749–759. URL: http://link.aps.org/doi/10.1103/PhysRev.40.749.

Wing, G. M. *A Primer on Integral Equations of the First Kind.* Society for Industrial Mathematics, 1987.

— *An Introduction to Transport Theory.* New York: Wiley, 1962.

Wolf, E. "Coherence and Radiometry". In: *J. Opt. Soc. Am.* 68 (1978), pp. 6–17. URL: http://www.opticsinfobase.org/abstract.cfm?URI=josa-68-1-6.

— *Introduction to the Theory of Coherence and Polarization of Light.* Cambridge University Press, Oct. 2007.

— "New Theory of Radiative Energy Transfer in Free Electromagnetic Fields". In: *Phys. Rev. D* 13 (1976), pp. 869–886.

— "The Development of Optical Coherence Theory". In: *Selected Works of Emil Wolf: with Commentary.* Ed. by E. Wolf. Singapore: World Scientific Publishing Company, 2001, pp. 620–633.

— "The influence of Young's interference experiment on the development of statistical optics". In: *Progress in Optics, Volume 50.* Ed. by E. Wolf. Vol. 50. Progress in Optics. Elsevier, 2007. Chap. 7, pp. 251–273. URL: http://www.sciencedirect.com/science/article/B7W5D-4R40N37-9/2/9f2d72e8dea2320f30b39310b1229a5c.

Wolf, K. B. *Geometric Optics on Phase Space.* Texts and Monographs in Physics. Springer, 2004.

Wunsch, J. *Microlocal Analysis and Evolution Equations.* Lecture notes from 2008 CMI/ETH Summer School on Evolution Equations. Dec. 2008. eprint: 0812.3181. URL: http://arxiv.org/abs/0812.3181.

Xiao, H., V. Rokhlin, and N. Yarvin. "Prolate spheroidal wavefunctions, quadrature and interpolation". In: *Inverse Problems* 17.4 (2001), pp. 805–838.

Yoshida, H. "Construction of higher order symplectic integrators". In: *Physics Letters A* 150.5-7 (Nov. 1990), pp. 262–268.

Zachos, C., D. Fairlie, and T. Curtright. *Quantum Mechanics in Phase Space: An Overview with Selected Papers*. World Scientific, 2005.

Zatz, H. R. "Galerkin Radiosity: A Higher Order Solution Method for Global Illumination". In: *International Conference on Computer Graphics and Interactive Techniques* (1993). URL: http://portal.acm.org/citation.cfm?id=166145.

Zayed, A. I. "A Generalization of the Prolate Spheroidal Wave Functions". In: *Proc. Amer. Math. Soc.* 135 (2007), pp. 2193–2203. URL: http://www.ams.org/journals/proc/2007-135-07/S0002-9939-07-08739-4/home.html.

Zhang, H., Y. Xu, and J. Zhang. "Reproducing Kernel Banach Spaces for Machine Learning". In: *The Journal of Machine Learning Research* 10 (Dec. 2009), pp. 2741–2775. URL: http://dl.acm.org/citation.cfm?id=1577069.1755878.

Zhang, H. and J. Zhang. "Frames, Riesz bases, and sampling expansions in Banach spaces via semi-inner products". In: *Applied and Computational Harmonic Analysis* 31.1 (July 2011), pp. 1–25. URL: http://dx.doi.org/10.1016/j.acha.2010.09.007.

Zhang, Z. and M. Levoy. "Wigner Distributions and How They Relate to the Light Field". In: *Proc. IEEE International Conference on Computational Photography*. IEEE Computer Society, 2009.

Zhao, Y., Y. Han, Z. Fan, F. Qiu, Y.-C. Kuo, A. E. Kaufman, and K. Mueller. "Visual simulation of heat shimmering and mirage." In: *IEEE transactions on visualization and computer graphics* 13.1 (Jan. 2007), pp. 179–89. URL: http://dl.acm.org/citation.cfm?id=1187626.1187771.

Zhong, G. and J. E. Marsden. "Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators". In: *Physics Letters A* 133.3 (Nov. 1988), pp. 134–139. URL: http://dx.doi.org/10.1016/0375-9601(88)90773-6.

Zhu, C., R. H. Byrd, P. Lu, and J. Nocedal. "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization". In: *ACM Transactions on Mathematical Software (TOMS)* 23.4 (1997). URL: http://portal.acm.org/citation.cfm?id=279236.

Zubairy, M. and E. Wolf. "Exact Equations for Radiative Transfer of Energy and Momentum in Free Electromagnetic Fields". In: *Optics Communications* 20.3 (1977), pp. 321–324. URL: http://dx.doi.org/10.1016/0030-4018(77)90192-4.

Index

Symbols

g-volume, 172

A

Abelian group, 28, 203
absorption coefficient, 5, 369
action functional, 229
ad operator, 218
ad* operator, 219
adjoint action, 216
    infinitesimal generator, 218
adjoint operator, 34
admissible chart, 109
aliasing, 414
alternation mapping, 152
analysis operator, 53
angular momentum of particle, 271
angular spectrum representation, 327
anti-self-adjoint operator, 48
approximation
    finite, 74
    optimal L_2, 55
approximation error, 75
approximation space, 75
area formulation of light transport, 448
atlas, 109
    compatibility, 110
atlas with boundary, 122
automorphism, 26

B

Banach space
    biorthogonal system, 43
    boundary, 122
    dimension, 42
    dual basis, 44
    half space, 122
    infinite dimensional, 42
    interior, 122
    isometry, 36
    operator norm, 37
    pairing, 39
    Schauder basis, 41
    strongly non-degenerate pairing, 40
    unconditional basis, 41
    weakly non-degenerate pairing, 40
basis
    analysis operator, 53
    coefficient functional, 43
    dual basis, 44, 51
    Gram matrix, 54
    orientation, 158
    reconstruction operator, 53
    synthesis operator, 53
basis constant, 42
basis matrix, 78
Beer-Lambert law, 369
Besov space, 175
Bessel bound, 51
Bessel sequence, 51
bidirectional path tracing, 6
bidirectional reflectance distribution functions, 368
bidirectional surface scattering distribution function, 367
biorthogonal basis, 51
biorthogonal system, 43
biorthogonality, 43
Bohr correspondence principle, 320
Boltzmann equation, 338, 339
boundary distance function, 353
bounded variation, 39
butterfly, 95

C

calculus of variations, 229
canonical 1-form, 238
canonical action, 266
canonical dual frame, 65
canonical mapping, 239
canonical momentum, 343
Cartan's formula, 181
Casimir function, 244
Casimir functions, 248
Cauchy sequence, 36
Cauchy-Green stress tensor, 192, 298
Cauchy-Schwarz inequality, 45

Nomenclature

$\bar{\Xi}_n$	normalized characteristic basis
(P, ω)	symplectic manifold
D^i	differential operator for the ith dimension
G/H	group quotient of G by H
G	Lie group (possibly infinite dimensional)
$G(q, \bar{q})$	geometry term for light transport
$H(q, p)$	Hamiltonian
H^s	Hilbert-Sobolev space
H_μ	reduced Hamiltonian at $\mu \in \mathfrak{g}^*$ for a Hamiltonian system with symmetry
J	momentum Hamiltonian
$L(\text{V}, \text{W})$	space of linear maps between V and W
$L(q, \dot{q})$	Lagrangian
$L^c(\mathbf{E}, \mathbf{F})$	space of continuous linear maps between the normed spaces $\mathbf{E}$ and $\mathbf{F}$.
P_μ	reduced phase space at $\mu \in \mathfrak{g}^*$
$S(q(t))$	action functional
$S^* Q$	cosphere bundle
S_k	group of permutations of k elements
$T\mathcal{M}$	tangent bundle
$T^* Q$	contangent bundle
$T_s^r(E)$	tensor bundle over the vector bundle E

$T^r_s(\mathcal{M})$	tensor bundle defined over the tangent bundle of the manifold $\mathcal{M}$
$T^r_s(\mathbf{E})$	space of (r, s) tensors on $\mathbf{E}$
W	Wigner transform
$W^{k,p}$	Sobolev space
X^μ_H	reduced Hamiltonian vector field at $\mu \in \mathfrak{g}^*$
$\mathrm{Den}(\mathcal{M})$	space of 1-densities on $\mathcal{M}$
$\mathrm{Diff}(\mathcal{M})$	Lie group of diffeomorphisms of $\mathcal{M}$
$\mathrm{Diff}_{\mathrm{can}}(P)$	Lie group of symplectic diffeomorphisms of the symplectic manifold P
$\mathrm{Diff}_\mu(\mathcal{M})$	Lie group of volume preserving diffeomorphisms of $\mathcal{M}$
$\mathcal{F}_0(\mathcal{M})$	functions homogeneous of degree one on $\mathcal{M}$
$\Gamma(E)$	space of sections of a vector bundle E
$\mathcal{H}_k(X)$	reproducing kernel Hilbert space
$\Lambda = \{\lambda_i\}^m_{i=1}$	reproducing point set for a reproducing kernel basis
$\mathcal{M}$	manifold
$\Omega(\mathcal{M})$	exterior algebra on $\mathcal{M}$
$\Omega^k(\mathcal{M})$	space of differential k-forms on $\mathcal{M}$
$\Omega^{p,q}(\mathcal{M})$	space of (p, q) double forms on $\mathcal{M}$
Ω_B	Paley-Wiener space of B-bandlimited functions in the Fourier domain
Q	configuration space of a mechanical system
$\mathbb{R}$	real numbers
$\mathbb{R}^+$	strictly positive real numbers
$\mathbb{R}^+_0$	positive real numbers including zero
V^+	algebraic dual space of the linear space V
$\|\cdot\|$	norm of a linear space
Ξ_n	characteristic basis
$\mathbb{Z}$	integers

$\bar{\ell}$	steady state energy density
$\bar{s}$	unit vector $\|\bar{s}\| = 1$
$\bigwedge(\mathbf{E})$	exterior algebra of $\mathbf{E}$
$\bigwedge^k(\mathbf{E})$	space of exterior k-forms over $\mathbf{E}$
$\mathrm{card}\,(X)$	cardinality of the set X
$\chi(P)$	indicator function for the set P
$\cong$	isomorphic
d	exterior derivative
δ	co-differential
$\ell(\gamma)$	length of the curve γ
ℓ_2^n	discrete Lebesgue space
ϵ	(di)electric permittivity
η	canonical transformation, $\eta \in \mathrm{Diff}_{\mathrm{can}}(P)$
η_t	curve $\eta_t : [a, b] \to \mathrm{Diff}_{\mathrm{can}}(P)$
$\mathfrak{s}$	Lie algebra of $\mathrm{Diff}_{\mathrm{can}}$
$\mathfrak{s}^*$	dual Lie algebra of $\mathrm{Diff}_{\mathrm{can}}$
$\hat{\mathcal{H}}_m(X)$	m-dimensional approximation space
$\hat{\mathrm{K}}_m$	operator matrix for a Hilbert-Schmidt operator
$\hat{f}_m$	m-dimensional approximation of the function f
$\hat{p}$	canonical momentum
$\ker(A)$	Kernel of the linear map A
$\langle , \rangle$	inner product
$\langle\!\langle , \rangle\!\rangle$	metric pairing between exterior forms
$\mathbb{F}$	fiber derivative
$\mathbb{J}$	symplectic matrix
$\mathbb{N}$	natural numbers
$\mathbb{N}_0$	natural numbers including zero
$\mathcal{A}$	analysis operator

$\mathcal{A}^*$	synthesis operator
$\mathcal{E}$	electromagnetic theory
	energy density
$\mathcal{F}(\mathcal{M})$	space of functions on a manifold
$\mathcal{R}$	reconstruction operator
$\mathcal{R}_\rho$	scattering operator
$\mathcal{S}$	frame operator
$\mathcal{T}$	scattering transport operator
$\mathcal{T}_s^r(\mathcal{M})$	space of section of the tangent tensor bundle
$\mathcal{U}_t$	light transport operator
$\mathfrak{X}(\mathcal{M})$	space of vector fields
$\mathfrak{X}^*(\mathcal{M})$	space of covector fields
$\mathfrak{X}_L(G)$	left invariant vector fields on a Lie group G
$\mathfrak{X}_R(G)$	right invariant vector fields on a Lie group G
$\mathfrak{X}_{\mathrm{Ham}}(P)$	space of Hamiltonian vector fields on the symplectic manifold P
$\mathrm{H}^k(\mathcal{M})$	k^{th} de Rahm cohomology group of $\mathcal{M}$
$\mathrm{Inv(g)}$	inversion map of a Lie group
$\mathrm{K}_\phi(\Lambda)$	reproducing matrix of a reproducing kernel basis defined over the reproducing points Λ with respect to the basis $\{\phi_i\}_{i=1}^n$
$\mathrm{f}(\Lambda)$	basis function coefficient vector of a reproducing kernel basis
$\mathbf{J}$	momentum map
$\mathbf{J}_L$	left Lie-Poisson momentum map
$\mathbf{J}_R$	right Lie-Poisson momentum map
μ	magnetic permeability
ω	symplectic 2-form
$\overline{\lim}$	limes superior
$\partial/\partial x^i$	local basis for $T\mathcal{M}$

$\partial \mathcal{M}$	boundary of a manifold
π_μ	quotient map for reduced phase space
σ_k^l	(k,l) shuffle
$\tilde{\ell}$	pre-radiance density
θ	canonical 1-form
$\varphi^* t$	pullback of a tensor field t
$\varphi_* t$	push-forward of a tensor field t
ϖ	Liouville form
$\vec{S}_{k+l}^k$	set of (k,l) shuffles
$\xi_\mathcal{M}$	infinitesimal generator of a Lie group action on the manifold $\mathcal{M}$
$\{,\}$	Poisson bracket
$d\bar{p}$	surface element of the sphere $d\bar{p} = \sin(\theta)\,d\theta d\phi$
dx^i	local basis for $T^*\mathcal{M}$
$g^n = \{g_n^{ij}\}$	natural co-metric for light transport
$g_n = \{g_{ij}^n\}$	natural metric for light transport
$k_y(x) = k(y,x)$	reproducing kernel
p	momentum
y_{lm}	Legendre spherical harmonic
i	imagniary unit
$S^+\mathcal{M}$	positive half space of the cosphere bundle $S^*\mathcal{M}$ as defined by the surface normal of the surface $\mathcal{M}$

''My work always tried to unite
 the truth with the beautiful,
 but when I had to choose
 one or the other,
 I usually chose the beautiful.''

 Hermann Weyl

''Wir müssen wissen,
 wir werden wissen.''

 David Hilbert