

PDEs on the Lie group $SE(3)$ and their application in diffusion-weighted MRI

Portegies, J.M.

Published: 12/03/2018

Document Version

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

Citation for published version (APA):

Portegies, J. M. (2018). PDEs on the Lie group $SE(3)$ and their application in diffusion-weighted MRI Eindhoven: Technische Universiteit Eindhoven

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**PDEs on the Lie Group $SE(3)$ and their Application in
Diffusion-Weighted MRI**

PDE: Partial Differential Equation

SE(3): Special Euclidean group of 3-dimensional rigid body motions

MRI: Magnetic Resonance Imaging

The work described in this thesis has received funding from the European Research Council under the European Community's 7th Framework Programme (FP7/2007-2014)/ERC grant agreement No. 335555 (Lie Analysis)



© Copyright Jorg Portegies, 2018

A catalogue record is available from the Eindhoven University of Technology Library

ISBN: 978-94-6233-897-5

Printed by Gildeprint, Enschede



**PDEs on the Lie Group $SE(3)$ and their Application in
Diffusion-Weighted MRI**

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor Promoties,
in het openbaar te verdedigen
op maandag 12 maart 2018 om 11:00 uur

door

Jorg Matthias Portegies

geboren te Middelburg

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter: prof.dr.ir. O.J. Boxma
promotor: prof.dr. L.M.J. Florack
copromotor: dr.ir. R. Duits
leden: prof.dr. K. Siddiqi (McGill University)
prof.dr. G. Citti (Università di Bologna)
prof.dr. R. Deriche (Université Côte d'Azur)
dr. J. Mirebeau (Université Paris-Sud)
prof.dr.ir. B.M. ter Haar Romeny
prof.dr.ir. B. Koren

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Summary

PDEs on the Lie Group $SE(3)$ and their Application in Diffusion-Weighted MRI

Diffusion-weighted Magnetic Resonance Imaging (dMRI) is often described as the unique non-invasive imaging method for obtaining insight in the structure of brain tissue. Due to this monopoly and thanks to its clinical feasibility, dMRI has attracted countless researchers, including myself, in all aspects of the technique. This thesis is the aggregation of my contributions to the dMRI field of research.

An important objective in this field is to obtain a detailed picture of the complete infrastructure that is formed by the neural pathways in the brain. For this picture both the neural fiber tract anatomy and the brain connectivity are of interest. However, there is a mismatch in scale between the resolution of the imaging modality and the typical size of the micro-structure that we want to observe: the typical voxel (3D pixel) scale is 1-2 mm, whereas neuronal axons typically have a diameter of the order of microns. Thanks to the structured packing of the axons in larger white matter tracts, they still have a noticeable influence on the amount of diffusion on the acquisition scale. Nevertheless, the indirect observation of the micro-structure through its larger-scale behavior makes it challenging to draw the right conclusions on the fiber anatomy and brain connectivity. Validation of the processing methods is not easy, since ground truth information is not at hand in most cases. Furthermore, acquisition noise, partial volume effects and various possible artifacts make correct interpretation of the data non-trivial. In this thesis I try to tackle some of these issues using methods that have a solid mathematical foundation.

A common feature of all chapters in this thesis, is that they deal with functions with position- and orientation-valued argument, which is (after some processing) the domain

of dMRI data. This exploits the fundamental property that the white matter is structured along elongated paths called fiber bundles, and the fact that the measured signal in dMRI is orientation-sensitive. Moreover, the concept of alignment of elements with both a position and an orientation generates a coupling between the two, influencing the geometry of the considered problems. Having this position-orientation domain as one of its main ingredients, this thesis strongly relies on the work on orientation scores started by Dr. Duits around 2005. In this line of work, grayscale images (originally 2D, but more recently also 3D) with primarily elongated features are lifted to a higher-dimensional position-orientation domain, via an orientation score transform that uses orientation-sensitive wavelets. The orientation score is not used here, because this is in some sense captured by the imaging modality, but both the rationale and the theory behind the type of processing of such orientation scores has formed the basis for much of the work in this thesis.

This thesis is divided into three parts. The introductory part (part I) consists of a more elaborate description of the objectives of the work, a brief discussion of the principles of dMRI and a motivation for the choice of mathematical theory and methods in Chapter 1. Since this thesis relies heavily on PDE theory, Lie groups (in particular the roto-translation group $SE(3)$) and optimal path methods, some basic concepts regarding these topics are described in Chapter 2.

Even though this thesis is strongly motivated from the application and in fact was started on the applied side of the spectrum, some mathematically interesting problems that emerged along the way have drawn some of the attention away from the application, towards the more theoretical work of this thesis. Although theory and application are closely connected and not at all separable, I have organized this thesis in such a way that part II mostly consists of theoretical contributions. Part III is reserved for the practical contributions.

In Chapter 3 I present a new approach to finding exact solution expressions to pure (hypo-elliptic) diffusion and convection-diffusion equations on the group $SE(3)$ of 3D translations and rotations. The 2D variants of these PDEs have previously been characterized as PDEs for contour enhancement and contour completion, respectively. This contour completion process was originally linked to the field of computer vision by David Mumford, and now both processes have been proven useful in 2D image processing applications. Similarly, in 3D they can be seen as PDEs for ‘bundle enhancement’ and ‘bundle completion’ and can thereby be valuable for processing of dMRI data. In order to solve the PDEs, finite difference methods can be used, or approximations of the solution kernel of the PDE. In this thesis, we combine classical, geometrical and functional analysis to derive new exact formulas for the solution kernels of both the time-dependent

and time-integrated case.

In Chapter 4, we propose a new approximation of the diffusion kernel on position-orientation-space, that has the symmetries that are inherent to the PDE. This easy-to-use approximation can be used for fast regularization of dMRI data, as we show in experiments later in the thesis. We also describe an alternative, stochastic viewpoint on these PDEs. The PDEs are in fact Fokker-Planck equations for the evolution of a probability density function. This stochastic viewpoint provides an intuitive interpretation of the PDEs in terms of random walks, and an alternative way to numerically obtain an approximating kernel.

Chapter 5 consists of a method for finding optimal paths through position-orientation data. With the appropriate optimization problem, these optimal paths can provide valuable information for the fiber tract anatomy, but also for brain connectivity measures. Two different models are considered to describe the optimization functionals, one that intuitively can be understood as a model for the optimal path of a car, and one that can be understood as a model for a car that has no reverse gear. These models have led to new theoretical results (in the form of a number of theorems), and they have required new numerical fast marching schemes.

Part III of this thesis is devoted to the practical results of the research. In Chapter 6, we show how the previously mentioned diffusion equation has advantages for tractography by using it in two places in the data processing pipeline. When the data is pre-processed such that a fiber orientation distribution is formed, this can be used as initial condition in the diffusion equation. This has a regularizing effect and enhances elongated structures in the data, while preserving crossings. Secondly, it can be used for post-processing of tractography results, in particular when probabilistic tractography is used. The impulse response to the diffusion process can be used in such a way that it provides a measure for how well aligned certain fibers are, and hence a method to remove spurious fibers from the data.

Chapter 7 continues the line of work of Chapter 5, focusing on the application. We show how the dMRI data can be used to impose a cost for moving, with a lower cost for moving along the direction of the fibers. Together with the intrinsic curvature penalization induced by the model, the paths that are solutions to the optimization problem seem to reflect well the brain anatomy and provide valuable connectivity information.

The final applied result, explained in Chapter 8, is an image reconstruction method: a framework in which multiple techniques are combined such as super-resolution, regularization of the image data or orientation distribution functions derived from it, and data-consistency. In proof-of-concept experiments, we were able to show advantages compared with standard image reconstruction methods.

This thesis is largely based on work that was done in close collaboration with a number of colleagues. In particular, I would like to acknowledge Jean-Marie Mirebeau for his great efforts and contributions concerning the fast marching paper that has formed the basis for Chapter 5. I thank Guido Janssen for his thorough proofreading of the theory in Chapter 3. I would like to thank Rutger Fick and Stephan Meesters for their valuable additions to Chapter 6, Pauly Ossenblok for her help with acquiring the dMRI data for the experiments in Chapter 6, and Vladimir Golkov for his input and enthusiasm during our work for Chapter 8. I thank Erik Bekkers for the collaboration on the work that has led to Chapter 4, and Gonzalo Sanguinetti for his guidance during the first part of my PhD, his support for the experiments and writing for the work in Chapter 6, and for being at the basis of the research that has led to the optimal path approach in Chapter 5. Finally, I would like to thank Remco Duits: as my supervisor and collaborator for all of my scientific output, you have been of inestimable value for this booklet.

Jorg Portegies
Eindhoven, 9th February 2018

Contents

Summary	v
I Introduction	1
1 Introduction	3
1.1 Basic principles of diffusion-weighted MRI	6
1.1.1 DMRI Reconstruction Methods	7
1.1.2 Tractography in dMRI	10
1.1.3 Brain anatomy and connectivity in dMRI	11
1.1.4 The dMRI pipeline	11
1.2 PDE methods in 2D image analysis	12
1.2.1 Image restoration	12
1.2.2 Image segmentation and optimal paths	15
1.3 PDE methods on position-orientation domain	15
1.3.1 Orientation scores	16
1.3.2 DMRI data as 3D orientation score	16
1.4 Contributions and outline	17
1.4.1 Overview Part II - Theoretical contributions	17
1.4.2 Overview Part III - Application to dMRI	18
2 Lie Groups and PDEs	19
2.1 Lie group $SE(3)$	20
2.1.1 The embedding of $\mathbb{R}^3 \times S^2$ into $SE(3)$	20
2.1.2 Lie algebras and differential operators	21
2.2 (Convection)-diffusion on $SE(3)$ and $\mathbb{R}^3 \times S^2$	24

2.2.1	Left-invariant convection-diffusion operators on $SE(3)$	24
2.2.2	Legal diffusion and convection-diffusion operators on $\mathbb{R}^3 \times S^2$	24
2.2.3	Relation with stochastic processes	26
2.3	The static Hamilton-Jacobi equation	27
2.3.1	Viscosity solutions	27
2.3.2	The shortest path optimization problem	28
2.3.3	Fast marching methods	29
2.3.4	Motivation for wavefront propagation on positions and orientations	33
II Theoretical contributions		35
3	Solutions of (Convection-)Diffusion Kernels on $SE(3)$	37
3.1	Introduction	38
3.1.1	Legal (convection-)diffusion operators on $\mathbb{R}^3 \times S^2$	38
3.1.2	Contributions and outline	40
3.2	Derivation of the exact solutions for hypo-elliptic diffusion on $\mathbb{R}^3 \times S^2$	41
3.2.1	Frequency-dependent choice of spherical variables	44
3.2.2	Separation of variables	45
3.2.3	Spheroidal wave equation	46
3.2.4	Sturm-Liouville form	48
3.2.5	Main theorem	48
3.2.6	Time-integrated processes and resolvent kernels	49
3.2.7	From the hypo-elliptic diffusion kernel to the elliptic diffusion kernel	50
3.3	Derivation of the exact solutions for convection-diffusion kernels on $\mathbb{R}^3 \times S^2$	51
3.3.1	The generalized spheroidal wave equation	53
3.3.2	Analysis of the time-integrated process	55
3.4	Matrix representation of the evolution and resolvent in a Fourier basis	60
3.4.1	Spherical harmonic expansions of the solutions to (convection-)diffusion equations	61
3.4.2	Numerical implementation of the discrete spherical transform	65
3.5	Conclusion	69
4	Approximations to Diffusion Kernels on $SE(3)$	71
4.1	Introduction	72
4.1.1	Contributions and outline	74

4.2	Symmetries of the diffusion kernel	74
4.3	Nilpotent approximation on SE(3) and sub-Riemannian distances	77
4.3.1	Approximation of the sub-Riemannian distance	80
4.4	New vs. previous kernel approximations	81
4.5	Stochastic approximation of diffusion kernels	84
4.6	Conclusion	86
5	Optimal Paths in SE(2) and SE(3)	89
5.1	Introduction	90
5.1.1	A distance function and shortest paths on $\mathbb{R}^d \times S^{d-1}$	91
5.1.2	Geometry of the Reeds-Shepp model	93
5.1.3	The eikonal equation and the fast marching algorithm	94
5.1.4	Contributions and outline	94
5.2	Main results	96
5.2.1	Preliminaries on the (quasi-)distance function and underlying geometry	96
5.2.2	Controllability of the Reeds-Shepp model	97
5.2.3	A continuous approximation for the Reeds-Shepp geometry	98
5.2.4	Points of interest in spatial projections of geodesics for the uniform cost case: cusps vs. keypoints	100
5.2.5	The Eikonal PDE formalism	103
5.3	Controllability properties: proof of Theorem 5.2.2, and Maxwell-points .	108
5.4	Cusps and keypoints: proof of Theorem 5.2.10	111
5.5	Eikonal equations and backtracking: proofs	113
5.6	Discretization of the Eikonal PDEs	116
5.6.1	Causal operators and the fast marching algorithm	116
5.7	Conclusion	120
III	Application to dMRI	123
6	Contextual PDE Flow for Improving Fiber Alignment	125
6.1	Introduction	126
6.1.1	Contributions and outline	128
6.2	Methods	129
6.2.1	A brief review of CSD	130
6.2.2	Contour enhancement (step A)	131
6.2.3	Tractography	132
6.2.4	Coherence quantification of fiber bundles (Step B)	134

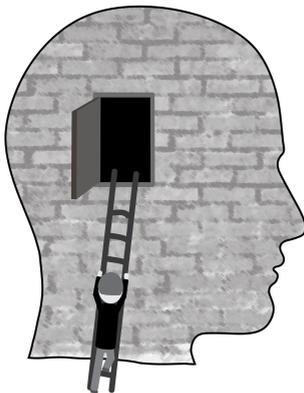
6.3	Experiments and results	137
6.3.1	HARDI Reconstruction Challenge	138
6.3.2	Evaluation and comparison on dMRI data	143
6.3.3	Improved reconstruction of the Optic Radiation	147
6.4	Conclusion	155
7	Optimal Paths in DMRI	157
7.1	Introduction	158
7.1.1	The shortest path problem	159
7.1.2	Overview of experiments and results	161
7.2	Applications in 2D	162
7.2.1	Shortest path to the exit in Centre Pompidou	162
7.2.2	Vessel tracking in retinal images	163
7.3	Application to dMRI data	165
7.3.1	Construction of the cost function	165
7.3.2	Influence of model parameters	165
7.3.3	Robustness to neighboring structures	166
7.3.4	ISBI HARDI Challenge 2013 dataset	167
7.3.5	ISMRM Tractography challenge 2015 dataset	170
7.4	Conclusion	173
8	Holistic Image Reconstruction for Diffusion MRI	175
8.1	Introduction	176
8.2	Methods	176
8.2.1	Image acquisition model	176
8.2.2	Holistic reconstruction	177
8.3	Results	183
8.4	Conclusion	184
	Concluding Remarks	187
A	Appendix to Chapter 3	189
A.1	Expansion of spheroidal wave functions	190
A.2	Generalized spheroidal wave functions	191
A.3	Proof of Theorem 3.3.6	192
A.4	Equivalent solutions via the Fourier transform on $SE(3)$	193
B	Appendix to Chapter 5	199
B.1	Well-posedness and convergence of the Reeds-Shepp models	200
B.1.1	Closedness of controllable paths	200

<i>Contents</i>	xiii
B.1.2 Specialization to the Reeds-Shepp models	202
B.2 Backtracking of geodesics in $(\mathbb{M}, d_{\mathcal{F}})$	205
B.3 Characterization of cusps: Proof of Lemma 5.2.6	206
Bibliography	209
List of publications	233
Acknowledgements	235
About the author	239

Part I

Introduction

Introduction



Scientists are thought to have ‘a good head on their shoulders’ (‘een goed stel hersens’, in Dutch). It is however not until you try to program a computer to perform certain tasks, like learning a language, playing a game, or distinguishing a bird from a plane in an image, that you really learn to appreciate how well the head on your shoulders really works and how hard it is to make a computer perform equally well. It is therefore not surprising that for such tasks, scientists often try to mimic the brain. However, a good understanding of the brain is a big challenge on its own.

In this thesis we focus on using computer vision for diffusion-weighted magnetic resonance imaging (dMRI) to obtain insight in the structure of the brain. By structure of the brain, we actually mean the structure of the white matter, which can be thought of as a large clutter of microscopically thin electrical wires, that connect brain regions and allow communication between them through the use of electrical pulses. The figure on the bottom of Fig. 1.1, obtained with a method called ‘tractography’, gives an impression of this wiring.

DMRI is the only non-invasive imaging method that can result in such structural information. It has gained popularity since the seminal paper [LBBL⁺86] of Le Bihan, who linked the anisotropy of the diffusion to the structure of the white matter. Nowadays, dMRI is a widely used tool in neuroscience, while slowly making its way into clinical procedures.

Although unique of its kind, dMRI is an imaging technique that is far from perfect. It has a long acquisition time (typically 10-30 minutes) and it produces images that are subject to noise, suffer from acquisition artifacts, have a much lower resolution than the typical size of the structures in the brain that we try to observe, etc. Moreover, a dMRI image is higher-dimensional than the 2D or 3D images that most image processing methods are developed for: it contains information about the amount of diffusion in different positions and different orientations.

We try to tackle some of the problems of dMRI from three different angles. Firstly, a diffusion equation is used to remove noise from the images, to improve subsequent tractography methods. Secondly, a Hamilton-Jacobi PDE is used to find globally shortest paths with respect to a cost function that is derived from the dMRI data, as an alternative to noise-sensitive local tractography methods. Finally, an optimization problem is formulated, that can incorporate various types of regularization, to reconstruct images on a higher resolution than they were acquired (super-resolution). Importantly, we include in each of these methods the idea that the white matter fibers are elongated structures with (often) limited curvature, and therefore influence the dMRI signal of neighboring voxels in a similar way.

Our methods have their origin in well-established 2D image processing methods. We show their advantage in dMRI image processing, in the end to improve our understanding

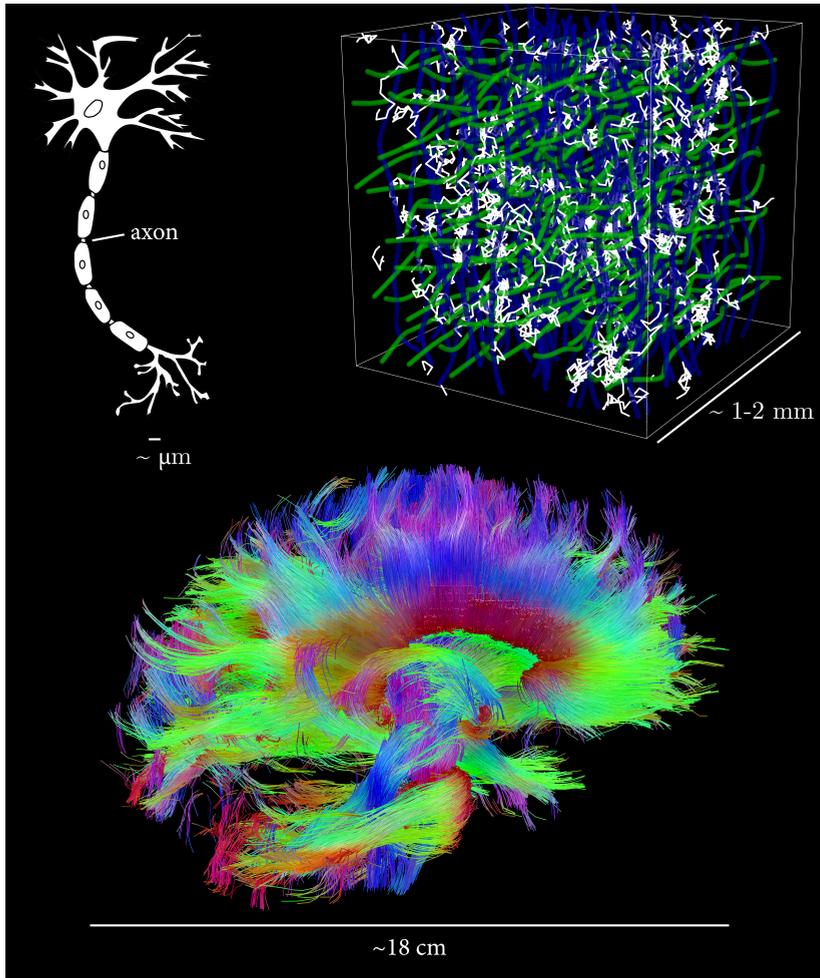


Figure 1.1: Top left: on the smallest scale, the cell level, we have neuronal axons that have a typical diameter of a couple of microns. Top right: The highly structured configuration of axons restricts the diffusion of water molecules in directions perpendicular to the structure. This leads to a measurable effect in the dMRI signal. This signal is measured with a spatial resolution of several millimeters. Bottom: finally, from the diffusion signal, we would like to obtain information on brain anatomy and connectivity on the largest scale (tractography image from <http://www.humanconnectomeproject.org/gallery/>).

of the structural connectivity in the brain white matter, adding to the body of knowledge that exists about the human brain. An overview of these methods, in very general terms, is given in Fig. 1.2. Before we outline the contributions of this thesis in more detail in Section 1.4, some background on the dMRI technique and computer vision methods is provided.

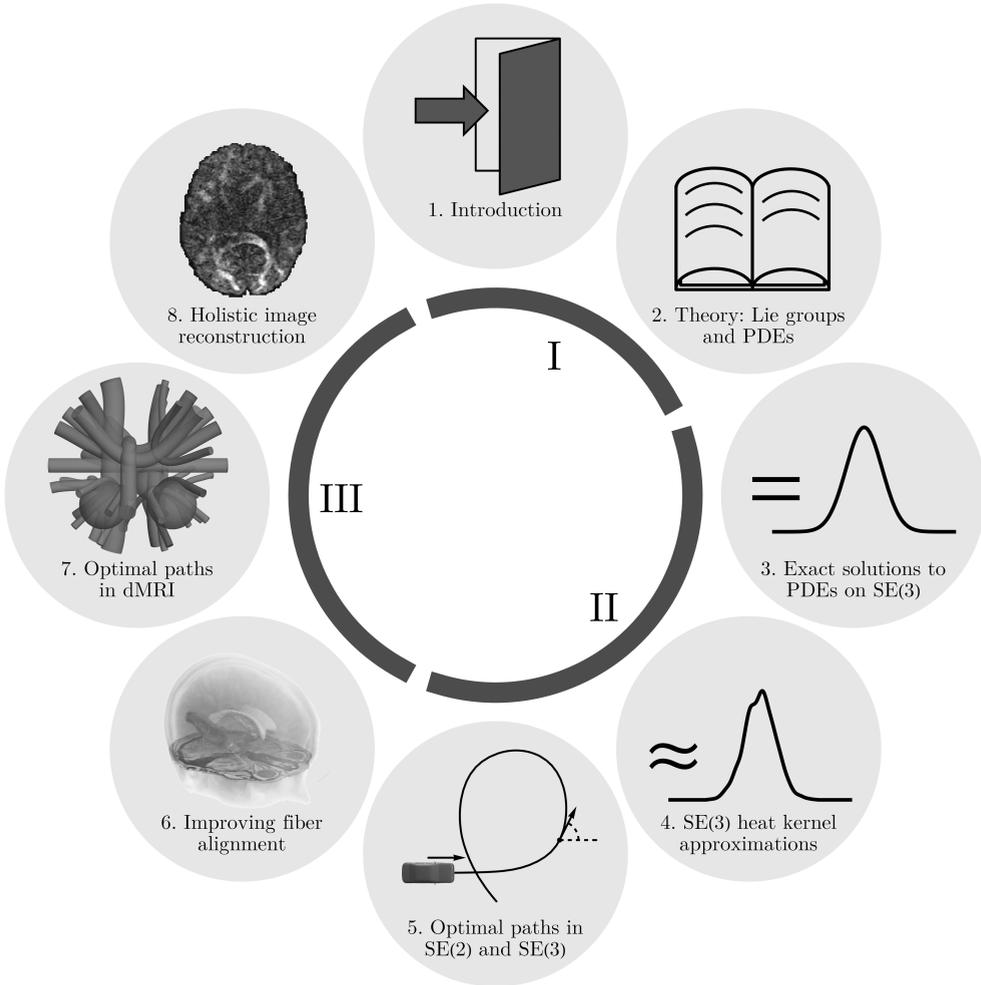


Figure 1.2: Overview of the parts and chapters of the thesis.

1.1 Basic principles of diffusion-weighted MRI

White and gray matter form the two most important parts of the brain. Gray matter mostly consists of neurons, the white matter mostly consists of axons, elongated projections of such neurons. They form the connections that can transmit information over long ranges between gray matter parts. The white matter thereby comprises the structural connectivity, or popularly speaking, the wiring of the brain. An individual axon has a diameter in the order of a micrometer, but can be several centimeters long. Axons are highly structured, forming larger fiber tracts that, because of their elongated structure,

restrict the diffusive motion of the water molecules in the direction perpendicular to the orientation of the tracts.

A dMRI scanner is sensitive to the amount of average diffusion of water molecules in different directions, measured within voxels ('volume-picture-element') of the size of 1-8 mm³. This gives insight in the underlying structure of the white matter, although the mismatch between the scanner resolution and the size of the microstructure makes it a challenging task to draw the right conclusions, see Fig. 1.1.

The imaging procedure is roughly as follows. A strong static magnetic field is used to orient hydrogen atoms in a uniform direction, the direction of the magnetic field. In fact, they spin around their axis in this direction with a frequency that depends on the magnetic field strength. A radio-frequency (RF) pulse is applied to bring these spins in the same phase and to bring some of these spins to a higher energy state. Subsequent relaxation results in an electromagnetic signal. Now a magnetic field gradient is applied for a fixed time, which speeds up the spins more where the field is stronger, causing the spins to dephase. Then, after again a short period of time, the same magnetic field gradient is applied, but with exactly the opposite direction (or a second RF pulse is applied to flip the spins 180 degrees, such that the same gradient in the same direction can be used). Ideally, this would bring all spins back into phase, but in reality, due to diffusion of the water molecules, the rephasing is imperfect, leading to a loss in signal. This loss in signal is exactly what gives the information on the amount of diffusion and thereby the underlying fiber structure: the higher the loss of signal, the more diffusion in the direction of the applied gradient field and the more likely that fibers are oriented in this direction. This described protocol is known as the Pulse-Gradient Spin-Echo (PGSE) sequence [ST65].

There is a number of parameters that influence the quality and level of detail in the images, that are usually combined to form the so-called b -value in s/mm². Apart from this b -value, there is also the choice of spatial and angular resolution, that has a large influence on the total scan time. The right choice of parameters should be made to highlight the desired features of the scan, with acceptable resolution and signal to noise ratio. After the data acquisition, several steps can be taken to improve the data, such as registration to correct motion artifacts [JBBS02], or Eddy current correction [JBP98]. In [DP12] a more complete overview of such steps is given.

1.1.1 DMRI Reconstruction Methods

The raw diffusion data from the scanner is known to be related to the diffusion propagator. This propagator describes in a voxel the probability that a water molecule has moved to a certain position after a certain time. So clearly a good estimation of the

propagator from the diffusion data provides valuable information about the restrictions the cell structures in a voxel put on the motion of water molecules.

One of the first and most widely used methods is Diffusion Tensor Imaging (DTI), proposed by Basser et al. [BMLB94] in 1994. It assumes that the propagator has the shape of a (3-variate) normal distribution, with a 3x3 covariance matrix, that represents the diffusion tensor. In each voxel, the six unknowns in this matrix are estimated from the diffusion data. Each tensor can be visualized as an ellipsoid, as illustrated in Fig. 1.3(b). In this visualization, the eigenvector of the matrix with the largest eigenvalue determines the principal direction of the ellipsoid is the eigenvector with the largest eigenvalue.

The DTI method has the advantage that it is relatively simple to apply, and that it works well on dMRI data with low angular resolution, that can be obtained with relatively short acquisition times. DTI has the limitation that it can only appropriately describe voxels in which there is only one predominant diffusion direction. In the work by Jeurissen et al. [JLT⁺13] two different models were used to find the prevalence of complex fiber structures, resulting in estimates that more than 60% of the white matter voxels consists of two or more fiber orientations.

In 2004, Tuch [Tuc04] presented a method called Q-ball imaging, that estimates in each voxel the orientation distribution function (ODF) that is the radial integral of the propagator. A slice of dMRI data, represented by ODFs obtained with Q-ball imaging, can be visualized as in Fig. 1.3(b). However, accurately estimating these ODFs requires High Angular Resolution Diffusion Imaging (HARDI) data, which as a longer acquisition time.

Glyph field visualization

Functions with position-orientation valued domain will appear throughout the entire thesis. In a dMRI image, they are sampled on a regular 3D Cartesian grid and on (more or less) uniform distributed orientations on the sphere. Such a function, say $U : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}^+$, can be visualized using glyphs. A glyph at a grid point $\mathbf{y} \in \mathbb{Z}^3$ is given by the surface $\{\mathbf{y} + \nu U(\mathbf{y}, \mathbf{n})\mathbf{n} \mid \mathbf{n} \in S^2\}$, for a suitable choice of $\nu \in \mathbb{R}$. Usually we choose $\nu > 0$ depending on the maximum of U , such that neighboring glyphs do not intersect. The coloring of each part of surface is orientation-dependent, according to a coloring standard in dMRI, see Fig. 1.3(a). The top-right figure shows an example of one such a glyph. Such a visualization of the entire brain is usually not very informative, so we often show just a slice.

In this thesis, we mostly use yet another model, called Constrained Spherical Deconvolution [TCGC04]. It does not estimate the ODF, but directly estimates the fiber orientation distribution (FOD) (apologies for the confusing acronyms). It calculates from

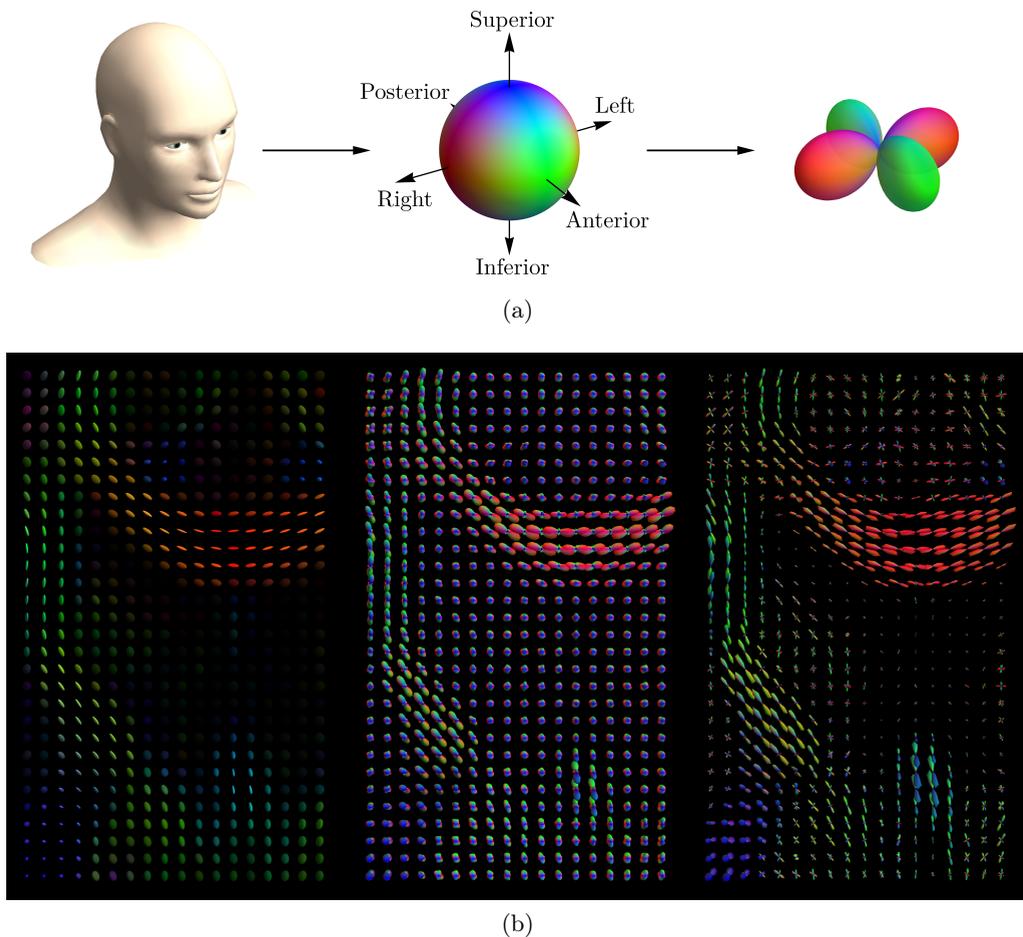


Figure 1.3: (a) Figure showing the coloring of orientations that has become standard in neuroimaging. For a voxel containing fibers going from both anterior-posterior and left-right, the fiber orientation distribution could look like the figure in the upper-right corner. (b) Example of a slice of data, processed with three well-established models in dMRI, from left to right: DTI, Q-ball and CSD.

the data the signal that would belong to a voxel with a single orientation of fibers, with which it deconvolves (decomposes) the signal in all other voxels. Hence it obtains an approximation of the distribution of fibers in each voxel.

The same data as for DTI and Q-ball imaging was processed using Constrained Spherical Deconvolution, and is visualized in Fig. 1.3(b). CSD has our preference, since it gives relatively sharp orientation profiles, making it quite suitable for subsequent tractography algorithms. However, with the right processing, e.g. using the Sharpening Deconvolution

Transform [DDKA09], our generic methods should work with ODFs obtained with Q-ball imaging as well.

The methods mentioned above are examples of single-shell methods, i.e., methods suitable for dMRI measurements done with a single b -value. More of these methods can be found in [DP12]. Other methods exist that rely on multi-shell data or even more intricate, that could give a more accurate representation of the white matter structure, at the cost of a longer acquisition. Examples of this are Diffusion Spectrum Imaging (DSI) [WHT⁺05], Diffusion Kurtosis Imaging [JHR⁺05] and multi-tissue multi-shell CSD [JTD⁺14a].

After modelling of the dMRI data using one of the above methods, a possible next step is to either extract some scalar measures from the reconstructed data, which allows for statistical comparisons between subjects. There exist a number of scalar measures that can be extracted either directly from the data or from one of the models discussed above. Well-known examples of such scalars are the apparent diffusion coefficient [ST65], fractional anisotropy [PB96] and generalized fractional anisotropy [Tuc04, OVM05]. A very popular alternative is to use a tractography method to get a more global view of the white matter structure.

1.1.2 Tractography in dMRI

Although the resolution of dMRI images is too low to directly estimate the white matter fibers, tractography provides a way of estimating the pathways of sufficiently thick, coherent bundles of such fibers, resulting in images such as the one on the bottom of Fig. 1.1. There is a vast number of different tractography methods and software tools; a detailed overview can be found in e.g. [TML11]. A recent comparison between many such methods can be found in [MNH⁺17]. Here we only name the different categories of tractography methods, together with their purpose and limitations.

Early methods were deterministic streamline methods, that simply followed from a seed point the main directions of the tensors obtained with DTI [MCCVZ99, BPP⁺00]. This has been extended to HARDI data, to improve reconstruction through voxels with complex fiber structures. Deterministic streamline methods are fast, even fast enough for real-time tractography [CWF⁺14], but they are the most sensitive to noise in the data, and often miss certain parts of data (false negatives).

Probabilistic models are less restrictive, in the sense that the direction in which the fibers are propagated is sampled from an orientation distribution based on the data [TCC12], and not just the local maximum in the data. It typically results in more true and false positives. Due to the probabilistic sampling, usually many more fibers need to be computed to obtain a sufficiently complete picture, which means longer computation and more post-processing to filter the false positives.

The last type that we mention here are geodesic methods, where truly shortest paths through the reconstructed dMRI data is computed. It provides in some sense the most probable connection between two points and immediately gives a distance that allows to extract certain connectivity measures. However, the optimization problem needs to be formulated carefully to obtain sensible paths and it is more complicated to implement. Geodesic methods play an important role in this thesis.

All methods, some more than others, suffer from noise originating from the scanner, acquisition artifacts and partial volume effects, that are likely to result in false fibers in the tractography output. To counter these (often local) reconstruction errors, we propose contextual regularization through the use of diffusion PDEs. In contextual image processing, we employ alignment and interaction of the glyphs that represent our fiber orientation distributions. Furthermore, we continue the line of work on geodesic tracking, based on Riemannian and Finslerian metrics, using new numerical methods for fast computation of optimal paths.

1.1.3 Brain anatomy and connectivity in dMRI

A correct picture of the brain anatomy using tractography has useful clinical applications; it can be used as an aid in presurgical planning, e.g. as a point of departure for efficient Electrical Stimulation Mapping. In another type of surgical operation called temporal lobe resection [FS63], for patients with a certain kind of epilepsy, it is very important to know the location of a particular white matter fiber bundle called the optic radiation.

Apart from these clinical benefits, it has become popular in neuroscience to make a complete map of neural connections in the brain into one ‘connectome’, the objective in for example the Human Connectome Project [VESB⁺13, SJX⁺13]. This can be done by dividing the cortex into subregions and summarizing the connections between regions that follow from tractography into a graph. How strongly they are connected could be extracted from the number of streamlines that are connecting them, or in the case of geodesic tracking, the geodesic length could provide such a measure. However, also in such a representation the imperfections of the tractography method can have a big influence on the accuracy of the obtained connectome.

1.1.4 The dMRI pipeline

As has become clear in the previous sections, processing dMRI from subject to end result is usually a concatenation of many different steps. This is often referred to as the dMRI pipeline, an example of which is depicted in Fig. 1.4. Although in this thesis we only consider neuroimages (brain scans), dMRI may also be used in clinical settings to scan e.g. the heart, other organs or muscles.

Inputs	Scanner	Data (pre-)processing	Data modeling	Further processing
Brain	Acquisition time	Brain extraction	DTI	Scalar maps (ADC, FA, ...)
Heart	Gradient pulse sequence	Segmentation	CSD	Regularization
Muscle	Resolution	Registration	Q-ball	Tractography
...	...	Image correction	DSI	Connectivity analysis
		Connectome

Figure 1.4: The dMRI processing pipeline divided into several steps, with for each step a number of central concepts.

Depending on the objective, different scanner parameters, pre-processing methods, models and further types of processing can be chosen. Please see e.g. [DP12, RTB⁺17, HJM⁺06] for in-depth overviews of many more subjects in the dMRI pipeline. The work in this thesis is focused on ‘further processing’ of the data, in particular regularization, tractography and connectivity analysis.

1.2 PDE methods in 2D image analysis

PDE-based methods have a significant share in the field of computer vision. Overviews of such methods in this field can be found in e.g. [CS05, Sap06, AK06]. The latter mentions three image processing/analysis categories in which PDEs play a major role: image restoration, segmentation and inpainting. To some extent, these three categories also appear in this thesis. We only treat the type of segmentation problems that makes use of optimal paths and we discuss this on an introductory level in Section 1.2.2. We first give a very brief overview of PDEs in image restoration and inpainting.

1.2.1 Image restoration

In this PDE formulation, a grayscale image is represented by a function $f : \Omega \rightarrow \mathbb{R}^+$, where $\Omega \in \mathbb{R}^2$ is the (typically rectangular) image domain. The PDEs that are used for various types of image restoration have the following general form:

$$\begin{cases} \partial_t u(\mathbf{x}, t) + F(\mathbf{x}, u(\mathbf{x}, t), \nabla u(\mathbf{x}, t), \nabla^2 u(\mathbf{x}, t)) = 0 & \text{in } (0, T) \times \Omega \\ \mathbf{n} \cdot \nabla u(\mathbf{x}, t) = 0 & \text{in } (0, T) \times \partial\Omega \\ u(\mathbf{x}, 0) = f(\mathbf{x}), & \text{in } \Omega. \end{cases} \quad (1.1)$$

Here \mathbf{n} indicates the normal at the boundary $\partial\Omega$, \mathbf{x} is the spatial variable (corresponding to a pixel).

The choice for the functional F will determine in what sense the solution $u(\mathbf{x}, T)$ of this equation at time T differs from the original image f . The image can for example be blurred, using (arguably) historically the most influential PDE, the heat equation

$$\begin{cases} \partial_t u(\mathbf{x}, t) = \Delta u(\mathbf{x}, t), \\ u(\mathbf{x}, 0) = f(\mathbf{x}), \end{cases} \quad (1.2)$$

with Δ the Laplace-operator. For each evolution time t (often interpreted as scale rather than time), the solution to this equation is a (Gaussian) blurred version of the image, see Fig. 1.5, resulting in a scale space of the original image. The blurred image corresponds to the convolution of the original image with a Gaussian kernel. It appears naturally from an axiomatic approach on how images should be processed, that has led to linear and non-linear scale space theory, to which many have contributed [Iij62, Wit83, Koe84, Flo97, Lin94, PGFM95, WII99, tHR03, DFGR04, FS04, DB07].

The standard heat equation typically causes too much blur around edges in an image. To counter this effect, Perona and Malik introduced their renowned adaptation [PM90]:

$$\begin{cases} \partial_t u(\mathbf{x}, t) = \nabla \cdot (g(|\nabla u|^2) \nabla u), \\ u(\mathbf{x}, 0) = f(\mathbf{x}). \end{cases} \quad (1.3)$$

Here $g : [0, \infty) \rightarrow (0, \infty]$ is typically a smooth, decreasing function, such that the amount of diffusion decreases when the image gradient $|\nabla u|$ becomes large, i.e., close to edges. This way surfaces receive smoothing similar to before, but (strong) edges are preserved much better. See Fig. 1.5(c) for an impression of the effect of such nonlinear diffusion. More on the axioms for nonlinear scale spaces can for example be found in [AGLM93].

Instead of changing just the amount of diffusion depending on certain local image features, it is also possible to make the diffusion direction-dependent. An example of this is the coherence-enhancing diffusion (CED) by Weickert [Wei99]. It makes use of the (smoothed) structure tensor of the (smoothed) image, that is known to be an edge detector, and has orthogonal eigenvectors along and perpendicular to edges. The corresponding nonlinear parabolic equation can be written as

$$\begin{cases} \partial_t u(\mathbf{x}, t) = \nabla \cdot (D(J_\rho(\nabla u_\sigma)) \nabla u), \\ u(\mathbf{x}, 0) = f(\mathbf{x}). \end{cases} \quad (1.4)$$

where J_ρ is the structure tensor, smoothed with a Gaussian kernel with standard deviation ρ . It depends on the gradient of the image u , also smoothed with a Gaussian kernel with standard deviation σ , to ensure elegant well-posedness results [Wei98]. See Fig. 1.5(d) for the effect of coherence-enhancing diffusion on an image of a painting of Van Gogh. This image was originally used already in Weickert's paper [Wei99].

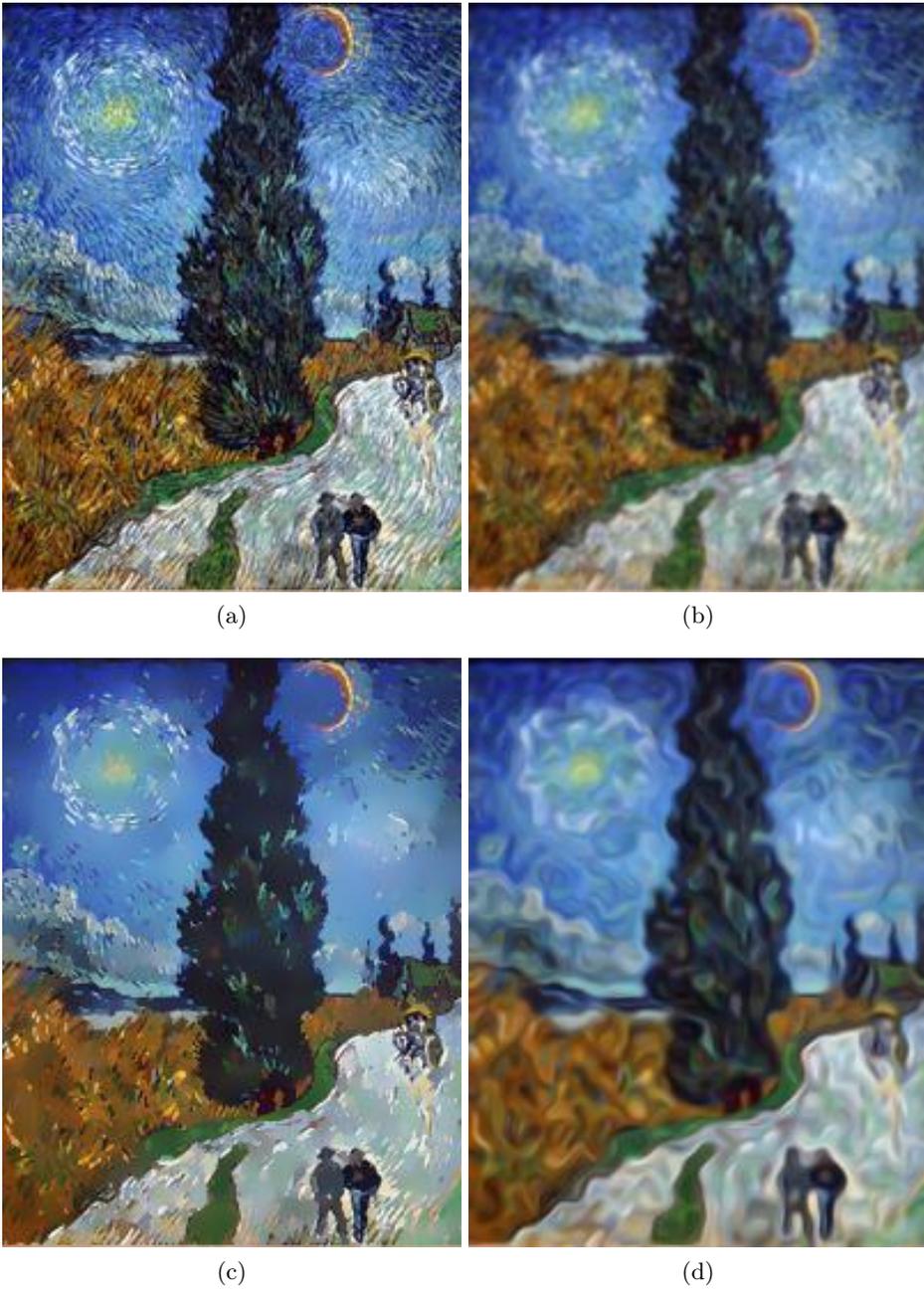


Figure 1.5: Van Gogh's 'Road with Cypress and Star' in 4 ways. (a) image of original painting. (b) Gaussian diffusion (c) Perona-Malik Diffusion (d) Coherence-enhancing diffusion.

1.2.2 Image segmentation and optimal paths

Another type of PDEs that is useful in computer vision is the Hamilton-Jacobi equation, a first order, nonlinear PDE [Eva10]. It can describe the motion of the boundary of a subset of the image domain, where properties of the image can be used to induce such motion. A possible application is image segmentation, where the boundary propagates (with a speed depending on its location in the image) to eventually form a closed curve around a shape or multiple shapes in the image, for example in geodesic active contour methods [CKS97, CK97]. This relies on the Hamilton-Jacobi equation of evolutionary type. An extensive review of this equation, numerical details and a range of applications can be found in [Set99].

We rather focus on the static Hamilton-Jacobi equation, which is the preferred choice when the speed of the front is positive and varies a lot between positions [Set99]. This is often the case in distance-to-origin type problems, i.e., when searching for the shortest connection to a seed point in an image, where e.g. the grayvalue of the image influences the cost. The distance map that is found by solving the Hamilton-Jacobi equation numerically, can be used to find the optimal path from each point in the image back to the origin by using a gradient descent.

The static Hamilton-Jacobi equation has the following form:

$$\begin{cases} H(\mathbf{x}, u(\mathbf{x}), \nabla u(\mathbf{x})) = 1, & \mathbf{x} \in \Omega \\ u(\mathbf{x}_s) = 0, \end{cases} \quad (1.5)$$

for a seed point $\mathbf{x}_s \in \Omega$. The Hamilton-Jacobi equation in general does not have a solution in the classical sense, but often has many weak solutions. The notion of viscosity solutions was introduced by Crandall and Lions [CL83, CIL92], that under certain conditions for H is the right concept to enforce uniqueness of solutions.

A very efficient method for solving the static Hamilton-Jacobi equation is the Fast Marching algorithm. Based on certain optimality principles, it constructs the solution with a single pass of all domain points. A number of applications that can be solved with Fast Marching can be found in [PPKC10]. However, whether the Fast Marching algorithm produces a solution that is a good approximation of the viscosity solution of the Hamilton-Jacobi equation, depends on a number of properties of the discretization. To be able to use the same theory in dMRI, the method requires a non-trivial extension to position-orientation domain, which forms an important part of this thesis.

1.3 PDE methods on position-orientation domain

In medical images, it is often the enhancement, segmentation or detection of elongated structures that is important. When we take retinal images as an example, the enhance-

ment of the blood vessels requires anisotropic diffusion, but this becomes problematic when they form complex structures, such as crossings and bifurcations. Instead of detecting these points and treating them differently, the images can alternatively be processed using orientation scores. Although in this thesis we do not make use of orientation score lifting, many of the methods in this thesis have had their origin in orientation scores.

1.3.1 Orientation scores

An orientation score of an image [Dui05] can be obtained using a wavelet transform, with an orientation sensitive filter. By computing the convolution of the images with rotated versions of the wavelet, a stack of responses of the image to the wavelet is obtained. When the image shows a structure that is in agreement with the orientation of the filter, the response will be high. This gives for each position of the image and every orientation of the filter a score, the domain of the orientation score is the space of positions and orientations $\mathbb{R}^2 \times S^1$. Because these positions and orientations can be identified with translations and rotations, Lie group theory and differential geometry on the group $SE(2)$ provides the right tools for processing of orientation scores.

Ideally, no information is lost when transforming back and forth between the image and the orientation score, which imposes certain design principles on the wavelets that are used. A specific type of (distributional) wavelet that allows such an invertible transform was first introduced by Kalitzin et al. [KRV99]. The stability of such a transform and the design of both non-distributional and distributional proper wavelets was then generalized by Duits [Dui05], leading to the theory of invertible orientation scores.

Yet another type of wavelet, the so-called cake-wavelet [Dui05], turned out to be more favorable, and has been used for image processing on many occasions since. The advantage of the orientation scores is that crossing lines in the original image are now on separate orientation-layers in the orientation score and cannot interfere with each other. This was exploited by Duits and Franken, leading to crossing-preserving evolutions of line structures in images [Fra08, FD09a, FD09b, DF10c, DF10b]. Image enhancement using orientation scores including a spatial scale parameter was proposed in [SD15]. A range of different applications of orientation scores in retinal image analysis can be found in [Bek17]. Also in the field of cortical modeling both linear and non-linear diffusions in position-orientation space have been studied [CS06, Pet03, BDGR12, PBG15, CFSS16].

1.3.2 DMRI data as 3D orientation score

Recently, 3D extensions of numeric and analytic versions of proper wavelets (such as the cake-wavelet) were developed [JDB15], resulting in 5D orientation scores with domain $\mathbb{R}^3 \times S^2$. As mentioned in Section 1.1, a large category of dMRI models aims to construct

from the raw data an orientation distribution, either to represent the diffusion profile, or directly the underlying fiber structure. This brings dMRI data naturally to a function on position-orientation domain, similar to the 3D orientation score, but without needing to apply a wavelet transform.

PDE-based processing on functions on $\mathbb{R}^3 \times S^2$ is still relatively unexplored, but a similar identification of positions with translations and orientations with rotations, allows us to use Lie group theory on the group $SE(3)$. A comprehensive study of (hypo-elliptic) PDEs in a general Lie group setting was presented in the influential paper of [RS76]. In a more applied setting, proof-of-concept experiments in [DF10a, DDHCG12, PAV⁺15] showed the potential of linear diffusion and morphological equations for enhancement of dMRI data. The main purpose in this thesis is therefore to introduce new methodology and carefully evaluate its benefit in dMRI, based on the same theoretical framework as the orientation score theory.

Processing of functions on positions-orientations comes at the cost of more complex and expensive numerics, in particular in 3D. A great effort in combining a large number of the numerical methods into a unifying Mathematica package was done by Frank Martin, see www.lieanalysis.nl, that was used for many of the results in this thesis.

1.4 Contributions and outline

In the previous sections, we have briefly touched upon the mathematical areas that this thesis relies on: Lie groups, PDEs, optimal path methods and fast marching. In Chapter 2, we will discuss them in some more detail. Together with this introduction, these two chapters form the first, introductory, part of the thesis. It forms the basis for the rest of the thesis, but mainly summarizes previous results. The new results obtained in the course of this research are presented in parts II and III of the thesis, part II being more focused on the theoretical results, and part III more on the practical applications. Figure 1.2 provides an overview of the parts and chapters of the thesis.

1.4.1 Overview Part II - Theoretical contributions

Chapter 3 describes an approach to find new exact solution expressions to diffusion and convection-diffusion equations on the group of 3D translations and rotations. To the best of our knowledge, no exact formulas were known for these PDEs, despite their wide use in image analysis and robotics and despite their attention in both mathematics and engineering. These solutions are expressed in terms of eigenfunctions of the operators that are obtained after a Fourier transform in the spatial variables. This new take on solving the equations also gives rise to a novel numerical solution.

In Chapter 4, we give two methods to approximate our exact diffusion kernel by more practical and simple formulas suited for fast evaluation. The first is an analytic expression, based on a nilpotent approximation of the Lie group $SE(3)$, greatly inspired by more general results obtained in [RS76,NSW85]. The second explores a stochastic point of view, since the PDEs are Fokker-Planck equations for the evolution of a probability density function. We use a Monte Carlo simulation to find another approximation of the kernel. We show for both approximations how well they approximate the exact solution.

Chapter 5 consists of an approach for finding globally optimal paths through position-orientation data in 2D and 3D. Essentially two different metrics are considered in the optimization problem, that are related to paths of cars with or without reverse gear. Optimal paths with respect to the model without reverse gear have the advantage that they do not have cusps, that are often unnatural in medical images. In a number of theorems and proofs, we carefully deal with theoretical questions on topics like the controllability of the system and convergence of optimization problems based on approximations of the original metrics. We also deal with numerical challenges that arise from using highly anisotropic metrics in Fast Marching methods.

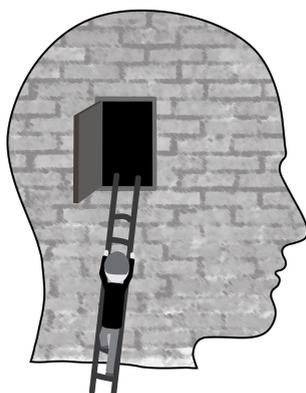
1.4.2 Overview Part III - Application to dMRI

The methods in the applied part of this thesis all fall within the last category (last column) of Fig. 1.4. In Chapter 6, we show how the previously mentioned diffusion equation has advantages for tractography in dMRI data by using it in two places in the data processing pipeline. We use constrained spherical deconvolution to construct a fiber orientation distribution as mentioned in Section 1.1, that we regularize with linear, restricted diffusion. Such contextual, crossing-preserving enhancement reduces the errors in the voxelwise reconstruction of CSD. The advantage for subsequent tractography is demonstrated on both synthetic and human brain dMRI data. Secondly, we introduce a measure to quantify how well-aligned tracked fibers are with respect to the entire tracking result, based on the diffusion kernel in position-orientation space. This gives a way to reduce the number of spurious fibers in probabilistic tractography results.

Chapter 7 shows how the theory on shortest path problems of Chapter 5 can be applied to dMRI images. We show how the dMRI data should be used as a cost function in the metric to obtain paths that are anatomically correct/plausible. We use a range of synthetic datasets with increasingly challenging and realistic bundle configurations to illustrate how connectivity measures can be derived from the fast marching results.

Finally, in Chapter 8 we show in proof-of-concept experiments that improved dMRI image reconstruction can be achieved, by combining several techniques such as super-resolution, FOD regularization, and data consistency in the acquisition space. We use an advanced primal-dual numerical method to solve the holistic optimization problem.

Lie Groups and PDEs



In this chapter we discuss the theoretical framework that forms the basis for the rest of the thesis. We start by giving some basic properties of the Lie group $SE(3)$, the corresponding Lie algebra and how Lie theory is used to define convection-diffusion equations on positions-orientation space. Secondly, we review the fundamentals of Hamilton-Jacobi equations and their relation to optimal control problems. Finally, we show how fast marching methods can be used to find their numerical solutions.

2.1 Lie group $SE(3)$

The group $SE(3)$ has several pseudonyms: the (special) euclidean motion group, the rigid-body motion group or the roto-translation group. A rigid body, which can not be deformed, can move by translation and rotation. Therefore, alternatively, the group can be written as $SE(3) = \mathbb{R}^3 \times SO(3)$. We can write each element as $g = (\mathbf{x}, \mathbf{R})$, such that $\mathbf{x} \in \mathbb{R}^3$ describes the translation and $\mathbf{R} \in SO(3)$ the rotation. The group product and inverse are defined by

$$gg' = (\mathbf{x}, \mathbf{R})(\mathbf{x}', \mathbf{R}') = (\mathbf{x} + \mathbf{R}\mathbf{x}', \mathbf{R}\mathbf{R}'), g^{-1} = (\mathbf{x}, \mathbf{R})^{-1} = (-\mathbf{R}^{-1}\mathbf{x}, \mathbf{R}^{-1}). \quad (2.1)$$

When using the matrix representation

$$g = (\mathbf{x}, \mathbf{R}) \sim \begin{pmatrix} \mathbf{R} & \mathbf{x} \\ \mathbf{0}^T & 1 \end{pmatrix} =: \mathbf{E}_g, \quad (2.2)$$

the group multiplication can conveniently be written as a matrix multiplication.

The group $SE(3)$ is a Lie group, meaning that it is a topological group, with the structure of a smooth manifold, such that multiplication and inversion are smooth [Kna02].

2.1.1 The embedding of $\mathbb{R}^3 \times S^2$ into $SE(3)$

Throughout this thesis, we mainly consider functions that have position-orientation-valued argument and map onto the real numbers, i.e. functions $U : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}$. Although there is an obvious resemblance between translations/rotations with positions/orientations, this needs to be handled with care. The difficulty lies in the fact that the sphere S^2 is a 2-parameter space, whereas the rotation group $SO(3)$ is a three-parameter group. To be able to use tools for (real-valued) functions on Lie groups, we need to make an identification between elements and equivalence classes as follows:

$$\mathbb{R}^3 \times S^2 \ni (\mathbf{x}, \mathbf{n}) \leftrightarrow \{(\mathbf{x}, \mathbf{R}_\mathbf{n}) \in SE(3) \mid \mathbf{R}_\mathbf{n}\mathbf{e}_z = \mathbf{n}\}.$$

Here $\mathbf{e}_z = (0, 0, 1)^T$ is a reference axis. If $\mathbf{R}_\mathbf{n}$ is any rotation such that $\mathbf{R}_\mathbf{n}\mathbf{e}_z = \mathbf{n}$, then clearly the following equivalence holds:

$$\mathbf{R}_n \sim \mathbf{R}_n \mathbf{R}_{\mathbf{e}_z, \alpha}, \quad \alpha \in (0, 2\pi].$$

We use the notation $\mathbf{R}_{\mathbf{a}, \phi}$ to write a counter-clockwise rotation around axis \mathbf{a} by angle ϕ . The group action \odot of $g \in SE(3)$ onto $(\mathbf{y}, \mathbf{n}) \in \mathbb{R}^3 \times S^2$ is defined by

$$g \odot (\mathbf{y}, \mathbf{n}) = (\mathbf{x}, \mathbf{R}) \odot (\mathbf{y}, \mathbf{n}) := (\mathbf{x} + \mathbf{R}\mathbf{y}, \mathbf{R}\mathbf{n}). \quad (2.3)$$

With this notation, we can say that elements $g, g' \in SE(3)$ are equivalent when:

$$g' \sim g \Leftrightarrow g' \odot (\mathbf{0}, \mathbf{e}_z) = g \odot (\mathbf{0}, \mathbf{e}_z).$$

Finally, we can define the coupled space of positions and orientations as the following Lie group quotient:

$$\mathbb{R}^3 \rtimes S^2 := SE(3) / (\{\mathbf{0}\} \times SO(2)). \quad (2.4)$$

We abuse the semi-direct-product notation \rtimes , usually reserved for the semi-direct product of groups, to emphasize the coupling between positions and orientations through the group action.

The embedding established above is used to relate functions $U : \mathbb{R}^3 \rtimes S^2 \rightarrow \mathbb{R}$ to functions $\tilde{U} : SE(3) \rightarrow \mathbb{R}$. We are restricted to functions that have the property that

$$U \leftrightarrow \tilde{U} \iff \left\{ U(\mathbf{x}, \mathbf{n}) = \tilde{U}(\mathbf{x}, \mathbf{R}_n) \text{ and } \tilde{U}(\mathbf{x}, \mathbf{R}) = U(\mathbf{x}, \mathbf{R}\mathbf{e}_z) \right\}. \quad (2.5)$$

This means that functions \tilde{U} derived as such carry a certain symmetry or invariance:

$$\tilde{U}(\mathbf{x}, \mathbf{R}) = \tilde{U}(\mathbf{x}, \mathbf{R}\mathbf{R}_{\mathbf{e}_z, \alpha}), \quad \alpha \in [0, 2\pi). \quad (2.6)$$

2.1.2 Lie algebras and differential operators

There is a Lie algebra associated to the group $SE(3)$, which is the tangent space $T_e(SE(3))$ of $SE(3)$ at the identity element. It consists of those elements that are, informally speaking, infinitesimally close to the identity. Using the matrix representation of the Lie group, matrix representations of elements of the Lie algebra are matrices \mathbf{M} such that

$$\exp \mathbf{M} = \begin{pmatrix} \mathbf{R} & \mathbf{x} \\ \mathbf{0}^T & 1 \end{pmatrix} = \mathbf{E}_g \sim (\mathbf{x}, \mathbf{R}) = g \in SE(3).$$

Here the map \exp is the matrix exponential, given by its power series. In general, also for groups that do not have a matrix representation, the exponential map maps elements from the algebra to the group. In this case, we can easily see that $\mathbf{R} = \exp \mathbf{X}$, with

$$\mathbf{R}^T = \mathbf{R}^{-1} \wedge \det \mathbf{R} = 1 \iff \exp(\mathbf{X})^T = \exp(\mathbf{X})^{-1} \iff \exp(\mathbf{X}^T) = \exp(-\mathbf{X}).$$

This implies \mathbf{M} is of the form $\begin{pmatrix} \mathbf{X} & \mathbf{x} \\ \mathbf{0} & 0 \end{pmatrix}$, with $\mathbf{X}^T + \mathbf{X} = 0$. Therefore the matrix Lie algebra is spanned by the basis

$$\begin{aligned} \mathbf{X}_1 &= \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & \mathbf{X}_2 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & \mathbf{X}_3 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ \mathbf{X}_4 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & \mathbf{X}_5 &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & \mathbf{X}_6 &= \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \end{aligned} \quad (2.7)$$

whereas the Lie algebra itself, i.e. $T_e(SE(3))$, is spanned by

$$\{A_1, \dots, A_6\} := \{\partial_x|_e, \partial_y|_e, \partial_z|_e, \partial_\gamma|_e, \partial_\beta|_e, \partial_\alpha|_e\}. \quad (2.8)$$

Remark 2.1.1. *As pointed out in [DF11], there are two coordinate charts needed for the rotational part, to be able to explicitly write down their formulas in terms of global coordinates. We mostly rely on the ZYZ-Euler angles, i.e.,*

$$\mathbf{R}_{\gamma,\beta,\alpha} = \mathbf{R}_{\mathbf{e}_z,\gamma} \mathbf{R}_{\mathbf{e}_y,\beta} \mathbf{R}_{\mathbf{e}_z,\alpha}.$$

Occasionally, we use XYZ-Euler angles, that we indicate with a tilde. They are needed to write down the basis of the Lie algebra as above, since the ZYZ coordinate chart is not defined at the identity element.

We have seen in Section 1.2 the importance of PDEs in 2D image analysis. To extend this theory to functions on $SE(3)$, we need differential operators for such functions. The derivative $X\tilde{U}$ (in direction \mathbf{X}) of a function \tilde{U} on the group can be written as

$$(X\tilde{U})(\mathbf{E}_g) := \left. \frac{d}{dt} \tilde{U}(\mathbf{E}_g \exp(t\mathbf{X})) \right|_{t=0}. \quad (2.9)$$

We choose this definition, using a perturbation on the right, because it commutes with the left-multiplication, i.e., these derivatives are left-invariant.

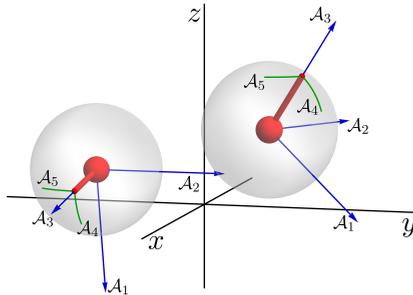


Figure 2.1: The left-invariant vector fields $\{\mathcal{A}_i\}_{i=1}^6$ can be seen as a moving derivative frame for functions on $SE(3)$. For two group elements $(\mathbf{x}, \mathbf{R}_n)$ we visualize the oriented particle (\mathbf{x}, \mathbf{n}) as a ball and stick. Then \mathcal{A}_3 is a spatial derivative in the direction \mathbf{n} , \mathcal{A}_1 and \mathcal{A}_2 are spatial derivatives orthogonal to this. \mathcal{A}_4 and \mathcal{A}_5 are derivatives along the green lines of the angular part of the function. The functions that we consider have an invariance that implies that $\mathcal{A}_6 = 0$.

Left-invariance is a requirement that is inherited from the orientation score theory, since translation- and rotation-invariance in position requires left-invariance in position-orientation space. Now each of the basis elements \mathbf{X}_i (and A_i) as defined above can be used to generate a left-invariant vector field using the push-forward of the left-multiplication. We denote this push-forward with $(L_g)_*$. Due to the semi-direct product structure, we hence obtain a moving derivative frame, that we define as

$$\mathcal{A}_i|_g \tilde{U} = \lim_{t \rightarrow 0} \frac{\tilde{U}(g e^{tA_i}) - \tilde{U}(g)}{t}, \quad i = 1, \dots, 6. \quad (2.10)$$

We write $\{\mathcal{A}_i\}_{i=1}^6$, where $\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$ are spatial and $\{\mathcal{A}_4, \mathcal{A}_5, \mathcal{A}_6\}$ are rotation vector fields, see Fig. 2.1. Note that (2.10) relates to (2.9) via the matrix representation. However, (2.10) only relies on the group product (2.1) and in principle does not require a matrix representation. The left-invariant vector fields are explicitly given by

$$\begin{aligned} \mathcal{A}_1|_g &= (\cos \alpha \cos \beta \cos \gamma - \sin \alpha \sin \gamma) \partial_x + (\sin \alpha \cos \gamma + \cos \alpha \cos \beta \sin \gamma) \partial_y - \cos \alpha \sin \beta \partial_z \\ \mathcal{A}_2|_g &= (-\sin \alpha \cos \beta \cos \gamma - \cos \alpha \sin \gamma) \partial_x + (\cos \alpha \cos \gamma - \sin \alpha \cos \beta \sin \gamma) \partial_y + \sin \alpha \sin \beta \partial_z \\ \mathcal{A}_3|_g &= \sin \beta \cos \gamma \partial_x + \sin \beta \sin \gamma \partial_y + \cos \beta \partial_z \\ \mathcal{A}_4|_g &= \cos \alpha \cot \beta \partial_\alpha + \sin \alpha \partial_\beta - \frac{\cos \alpha}{\sin \beta} \partial_\gamma, \\ \mathcal{A}_5|_g &= -\sin \alpha \cot \beta \partial_\alpha + \cos \alpha \partial_\beta + \frac{\sin \alpha}{\sin \beta} \partial_\gamma, \\ \mathcal{A}_6|_g &= \partial_\alpha, \end{aligned} \quad (2.11)$$

for $\beta \notin \{0, \pi\}$. Left-invariant vector fields expressed in the second coordinate chart can be found in e.g. [DF10a]. They form the basis for the differential operators in the PDEs that we introduce in the next section.

2.2 (Convection)-diffusion on $SE(3)$ and $\mathbb{R}^3 \times S^2$

2.2.1 Left-invariant convection-diffusion operators on $SE(3)$

Let $\tilde{U} : SE(3) \rightarrow \mathbb{R}^+$ be a square integrable function, the original ‘image’. The general evolution equation with \tilde{U} as initial condition is given by:

$$\begin{cases} \partial_t \tilde{W}(g, t) = \tilde{Q} \tilde{W}(g, t), & \text{for all } g \in SE(3), t \geq 0, \\ \tilde{W}(g, 0) = \tilde{U}(g), & \text{for all } g \in SE(3). \end{cases} \quad (2.12)$$

Here \tilde{Q} is the generator of the evolution on the group, where we restrict ourselves to generators such that the evolution becomes a linear, second order convection-diffusion process. Moreover, \tilde{Q} should be a left-invariant operator and is composed of the left-invariant differential operators $\underline{\mathcal{A}} = (\mathcal{A}_1, \dots, \mathcal{A}_6)$, recall their definition in Section 2.1.2. The general form of such operators is

$$\tilde{Q}^{\mathbf{D}, \mathbf{a}} = \sum_{i=1}^6 -a_i \mathcal{A}_i + \sum_{i,j=1}^6 \mathcal{A}_i D_{ij} \mathcal{A}_j. \quad (2.13)$$

Here $\mathbf{D} = [D_{ij}] \in \mathbb{R}^{6 \times 6}$ is a symmetric positive semi-definite 6×6 -matrix, and $\mathbf{a} = (a_i)_{i=1}^6 \in \mathbb{R}^6$. We use $\tilde{Y}_t(\tilde{U}) = \tilde{W}(\cdot, t)$ to denote the solution of the above evolution equation, so $\tilde{Y}_t : \mathbb{L}_2(SE(3)) \rightarrow \mathbb{L}_2(SE(3))$. The evolution can then formally be solved by convolution with the integrable solution kernel or impulse response $\tilde{K}_t : SE(3) \rightarrow \mathbb{R}^+$:

$$\tilde{Y}_t(\tilde{U})(g) = (e^{t\tilde{Q}^{\mathbf{D}, \mathbf{a}}} \tilde{U})(g) = (\tilde{K}_t *_{SE(3)} \tilde{U})(g) = \int_{SE(3)} \tilde{K}_t(h^{-1}g) \tilde{U}(h) dh, \quad (2.14)$$

where we use the Haar measure on $SE(3)$, which is the product of the Lebesgue measure on \mathbb{R}^3 and the Haar measure on $SO(3)$.

2.2.2 Legal diffusion and convection-diffusion operators on $\mathbb{R}^3 \times S^2$

From the dMRI application, we often encounter functions defined on $\mathbb{R}^3 \times S^2$ instead of the functions on $SE(3)$ as above. As argued in Section 2.1.1, a coupling between positions and orientations is required, which can be obtained by embedding the space $\mathbb{R}^3 \times S^2$ in $SE(3)$. For functions defined on the quotient $\mathbb{R}^3 \times S^2$ the same machinery as in the previous section can be used, although some restrictions apply. In [DDHCG12] it is shown that operators on \tilde{U} are legal (i.e., they correspond to well-defined operators on U that commute with rotations/translations) if and only if the following holds:

Definition 2.2.1. An operator $\tilde{\Phi} : \mathbb{L}_2(SE(3)) \rightarrow \mathbb{L}_2(SE(3))$ is called legal if

$$\begin{cases} \mathcal{L}_g \circ \tilde{\Phi} = \tilde{\Phi} \circ \mathcal{L}_g & \forall g \in SE(3), \\ \tilde{\Phi} \circ R_h = \tilde{\Phi} \circ R_{h'} \circ \tilde{\Phi} & \forall h = (\mathbf{0}, \mathbf{R}_{\mathbf{e}_z, \alpha}), h' = (\mathbf{0}, \mathbf{R}_{\mathbf{e}_z, \alpha'}), \alpha, \alpha' \in (0, 2\pi], \end{cases}$$

with \mathcal{L}_g and R_h the left- and right-regular action on $SE(3)$ respectively, see [DF11].

Here the left-regular action \mathcal{L}_g and right-regular action \mathcal{R}_g are defined as

$$(\mathcal{L}_g \tilde{U})(g') := \tilde{U}(g^{-1}g'), \quad (\mathcal{R}_g \tilde{U})(g') := \tilde{U}(g'g), \quad g, g' \in SE(3) \quad (2.15)$$

Legal operators $\tilde{\Phi}$ induce well-posed operators Φ on $\mathbb{L}_2(\mathbb{R}^3 \times S^2)$ via $\Phi(U)(\mathbf{y}, \mathbf{n}) = \tilde{\Phi}(\tilde{U})(\mathbf{y}, \mathbf{R}_\mathbf{n})$.

For the diffusion case, an equivalent statement is that the diffusion matrix \mathbf{D} is invariant under conjugation with $\mathbf{Z}_\alpha = \mathbf{R}_{\mathbf{e}_z, \alpha} \oplus \mathbf{R}_{\mathbf{e}_z, \alpha} \in SO(6)$, i.e. $\mathbf{Z}_\alpha^{-1} \mathbf{D} \mathbf{Z}_\alpha = \mathbf{D}$. Among the few legal, left-invariant diffusion and convection-diffusion generators on $\mathbb{R}^3 \times S^2$, recall (2.13), are the pure hypo-elliptic diffusion case and the hypo-elliptic convection-diffusion case, for details see [DF10a, DDHCG12]. We denote the evolution generators of these two cases on the group and on the quotient with \tilde{Q}_i and Q_i , $i = 1, 2$, respectively. They are defined as:

$$\tilde{Q}_1 := D_{33} \mathcal{A}_3^2 + (D_{44} \mathcal{A}_4^2 + D_{55} \mathcal{A}_5^2), \quad (2.16)$$

with $D_{33} > 0, D_{44} = D_{55} > 0$. The generator \tilde{Q}_1 acts on sufficiently smooth functions \tilde{U} and can be identified with a generator Q_1 acting on functions U :

$$Q_1 := D_{33} (\mathbf{n} \cdot \nabla_{\mathbb{R}^3})^2 + D_{44} \Delta_{S^2}. \quad (2.17)$$

Here $\nabla_{\mathbb{R}^3}$ denotes the gradient on \mathbb{R}^3 , and Δ_{S^2} is the Laplace-Beltrami operator on the sphere S^2 . Also the following (hypo-elliptic) convection-diffusion generator can be identified with a legal generator on $\mathbb{R}^3 \times S^2$:

$$\tilde{Q}_2 := -\mathcal{A}_3 + (D_{44} \mathcal{A}_4^2 + D_{55} \mathcal{A}_5^2), \quad (2.18)$$

again with $D_{44} = D_{55} > 0$. The corresponding generator Q_2 acting on sufficiently smooth functions U is defined as:

$$Q_2 := -(\mathbf{n} \cdot \nabla_{\mathbb{R}^3}) + D_{44} \Delta_{S^2}. \quad (2.19)$$

Remark 2.2.2. The way \tilde{Q}_i and Q_i act on functions \tilde{W} and W , respectively, is related by the following identities:

$$\begin{aligned} (\mathbf{n} \cdot \nabla_{\mathbb{R}^3} W)(\mathbf{y}, \mathbf{n}) &= (\mathcal{A}_3 \tilde{W})(\mathbf{y}, \mathbf{R}_n), \\ (\Delta_{S^2} W)(\mathbf{y}, \mathbf{n}) &= (\mathcal{A}_4^2 + \mathcal{A}_5^2) \tilde{W}(\mathbf{y}, \mathbf{R}_n). \end{aligned} \quad (2.20)$$

As a result, the following relation holds:

$$(\tilde{Q}_i \tilde{W})(\mathbf{y}, \mathbf{R}_n) = (Q_i W)(\mathbf{y}, \mathbf{n}), \quad i = 1, 2, \quad (2.21)$$

regardless of the choice of rotation for \mathbf{R}_n (mapping \mathbf{e}_z onto \mathbf{n}).

In summary, we focus on the following two PDE systems using (2.17) and (2.19):

$$\begin{cases} \partial_t W(\mathbf{y}, \mathbf{n}, t) = (Q_i W)(\mathbf{y}, \mathbf{n}, t), \\ W(\mathbf{y}, \mathbf{n}, 0) = U(\mathbf{y}, \mathbf{n}), \end{cases} \quad (2.22)$$

with $(\mathbf{y}, \mathbf{n}) \in \mathbb{R}^3 \times S^2$, $t \geq 0$, $i = 1, 2$. In Chapters 3 and 4, we want to find an exact solution and approximations for the impulse response or convolution kernel of these PDEs, i.e., the solution with initial condition $U(\mathbf{y}, \mathbf{n}) = \delta_{(\mathbf{0}, \mathbf{e}_z)}(\mathbf{y}, \mathbf{n})$. Solutions of (2.22) are found by $\mathbb{R}^3 \times S^2$ -convolution with the exact solution kernel $p_t : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}^+$:

$$\begin{aligned} W(\mathbf{y}, \mathbf{n}, t) &= (p_t *_{\mathbb{R}^3 \times S^2} U)(\mathbf{y}, \mathbf{n}) \\ &= \int_{S^2} \int_{\mathbb{R}^3} p_t(\mathbf{R}_n^T(\mathbf{y} - \mathbf{y}'), \mathbf{R}_n^T \mathbf{n}) \cdot U(\mathbf{y}', \mathbf{n}') d\mathbf{y}' d\sigma(\mathbf{n}'). \end{aligned} \quad (2.23)$$

In Chapter 6, we study the practicalities of using this type of diffusion on dMRI images.

2.2.3 Relation with stochastic processes

Any evolution equation corresponding to a generator $Q^{\mathbf{D}, \mathbf{a}}$ as above, is in fact a Fokker-Planck equation for the evolution of a probability density function. The (convection)-diffusion processes generated by Q_1 and Q_2 were called contour enhancement and contour completion in [DF10a], respectively, because of the stochastic processes they relate to. The convolution kernels of the PDEs can also be obtained with a Monte Carlo simulation by accumulating infinitely many random walks starting at $(\mathbf{0}, \mathbf{e}_z)$ that obey the underlying stochastic process and have traveling time t . The SE(2) process for contour completion is better known as Mumford's direction process [Mum94] and finds its application in computer vision. The relation with stochastic processes is further explored in Chapter 4. The reader could already have a look at Fig. 4.1 in that chapter for some intuition.

2.3 The static Hamilton-Jacobi equation

In Section 1.2.2, we have briefly introduced the static Hamilton-Jacobi equation

$$\begin{cases} H(\mathbf{x}, u(\mathbf{x}), \nabla u(\mathbf{x})) = 1, & \mathbf{x} \in \Omega, \\ u(\mathbf{x}_s) = 0, & \mathbf{x}_s \in \Omega, \end{cases} \quad (2.24)$$

with Ω some open subset of \mathbb{R}^n (or e.g. $\mathbb{R}^n \times S^{n-1}$ as in Chapter 5) and H the Hamiltonian (or rather, $H - 1$). The boundary condition also often appears as $u = \phi$ on the boundary of the domain $\partial\Omega$, but we will only consider the case with a single ‘source’. In this section we continue the discussion on this type of equations in more detail. We clarify the relation between the HJ equation and optimal paths, we discuss viscosity solutions and we provide some more insight in the fast marching method and its working principles.

2.3.1 Viscosity solutions

In many cases, the HJ equations do not admit classical solutions, and we need to resort to generalized solutions. On the other hand, the usual concept of weak solutions is too weak. Consider the simple 1D example where we want to find the distance to the origin, i.e., the solution should be $u(x) = |x|$. The corresponding eikonal equation is

$$|u'(x)| = 1, \quad u(0) = 0, \quad (2.25)$$

that does not have a classical solution, but infinitely many weak solutions. Ideally, we define a type of generalized solution that is unique, and coincides with the value function of the optimization problem corresponding to the HJ equation. This is the case for so-called viscosity solutions.

Definition 2.3.1 (Viscosity solution). *Let $u \in C(\Omega)$, then*

1. *u is a viscosity subsolution if for every function $\phi \in C^1(\Omega)$ such that $u - \phi$ has a local maximum at $\mathbf{x} \in \Omega$, that $H(\mathbf{x}, u(\mathbf{x}), \nabla\phi(\mathbf{x})) \leq 1$.*
2. *u is a viscosity supersolution if for every function $\phi \in C^1(\Omega)$ such that $u - \phi$ has a local minimum at $\mathbf{x} \in \Omega$, that $H(\mathbf{x}, u(\mathbf{x}), \nabla\phi(\mathbf{x})) \geq 1$.*

Then u is a viscosity solution if it is both a viscosity super- and subsolution.

It is not difficult to check that the function $u(x) = |x|$ is indeed a viscosity solution in the example above. The term viscosity solution comes from the fact that historically, a small viscous term was added to the equation to impose uniqueness of a solution, and then taking a limit to let the term vanish [Eva10]. The concept of viscosity solutions

above works in more general cases, but in cases where taking this limit works, it also coincides with the viscosity solution according to the formal definition. Also note that continuity of the solution is the only requirement.

Next, it can be shown that when H is equicontinuous in the spatial variable, that the boundary value problem admits at most one solution.

2.3.2 The shortest path optimization problem

We study the Hamilton-Jacobi equation because of its relation with shortest path optimization problems, in which we are actually interested. Here we will briefly give some background to this relation.

We use (as our most general case) a Finsler function $\mathcal{F} : T(\mathbb{M}) \rightarrow \mathbb{R}^+$ on the tangent space of a manifold \mathbb{M} to define the length of a path $\gamma : [0, 1] \rightarrow \mathbb{M}$:

$$\text{Length}_{\mathcal{F}}(\gamma) = \int_0^1 \mathcal{F}(\gamma(t), \dot{\gamma}(t)) dt. \quad (2.26)$$

The distance $d_{\mathcal{F}} : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{R}^+$ between two points $\mathbf{p}, \mathbf{q} \in \mathbb{M}$ is the length of the shortest (Lipschitz-continuous) path connecting those points:

$$d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) := \inf\{\text{Length}_{\mathcal{F}}(\gamma) \mid \gamma(0) = \mathbf{p}, \gamma(1) = \mathbf{q}, \gamma \in \text{Lip}([0, 1], \mathbb{M})\}. \quad (2.27)$$

This can be reformulated in terms of a control system, where the dynamics of the ‘particle’ subject to the control u is trivially described by

$$\dot{\gamma}(t) = u, \quad \text{Cost}_{\mathcal{F}}(u) = \int_0^1 \mathcal{F}(\gamma(t), u(t)) dt,$$

where u should be such that $\gamma(1) = \mathbf{q}$. The optimal control problem would be to find u within a predetermined control set that minimizes the cost.

We define $U : \mathbb{M} \rightarrow \mathbb{R}^+$ to be the value function of the distance with respect to some source point $\mathbf{p}_S \in \mathbb{M}$:

$$U(\mathbf{p}) := d_{\mathcal{F}}(\mathbf{p}_S, \mathbf{p}). \quad (2.28)$$

Then using Pontryagin’s and Bellman’s optimality principles, see e.g. [BCD97, AS04], it can be shown that this value function is the viscosity solution to the HJ equation:

$$\mathcal{F}^*(\mathbf{p}, dU(\mathbf{p})) = 1, \quad \mathbf{p} \in \mathbb{M} \setminus \{\mathbf{p}_S\},$$

with $U(\mathbf{p}_S) = 0$. Here \mathcal{F}^* is the dual of \mathcal{F} . Once the value function U is found, the optimal path (or the optimal control) can be derived from it:

$$\begin{cases} \dot{\gamma}(t) = L \, d_{\mathbf{p}} \mathcal{F}^*(\gamma(t), dU(\gamma(t))), & L := d_{\mathcal{F}}(\mathbf{p}_S, \mathbf{p}) \\ \gamma(0) = \mathbf{p}_S, & \gamma(1) = \mathbf{p}, \end{cases}$$

where $d_{\mathbf{p}}$ denotes differentiation w.r.t. the second entry of \mathcal{F}^* . This approach, using the Hamilton-Jacobi PDE framework (and its viscosity solution), became popular in image processing on \mathbb{R}^d since the pioneering work by Cohen and Kimmel [CK97].

There is a very intuitive meaning to this setup: level sets of the distance map $U(\mathbf{p}) = c$ move away from the source point when c increases, just like the propagation of a wavefront. The ODE with which the optimal path is obtained, is like a gradient descent on U .

2.3.3 Fast marching methods

Now that we have established the relation between optimal path problems and the Hamilton-Jacobi equation, the question becomes how to solve them. Fast marching methods provide an answer. They are well-known for their efficiency, since they build the solution in a *single pass* of the (discretized) domain, in contrast to methods that require iterative updates of the entire domain.

The Fast Marching algorithm for the computation of U defined as in (2.28), with source point \mathbf{p}_S , is roughly as follows:

Algorithm 1 Fast marching

Pre-compute the stencils

Tag the source point \mathbf{p}_S as Accepted, all points connected to \mathbf{p}_S as Trial, and other points as Far

While Trial not empty

Find a Trial point \mathbf{p} minimizing U

Tag \mathbf{p} as Accepted.

For each point \mathbf{q} connected to \mathbf{p}

If \mathbf{q} is Far, then tag it as Trial.

Update $U(\mathbf{q})$.

There are basically two challenges in this algorithm: the computation of stencils, that define the connectivity between points, and an efficient update procedure. The stencils should be designed in such a way that the algorithm is consistent, sufficiently accurate and efficient.

One of the first approaches to fast marching was introduced by Tsitsiklis [Tsi95], for shortest path problems with 2D isotropic metrics of the form:

$$\mathcal{F}(\mathbf{x}, \dot{\mathbf{x}}) = \mathcal{C}(\mathbf{x}) \|\dot{\mathbf{x}}\|, \quad \mathbf{x} \in \mathbb{M} = \mathbb{R}^2 \tag{2.29}$$

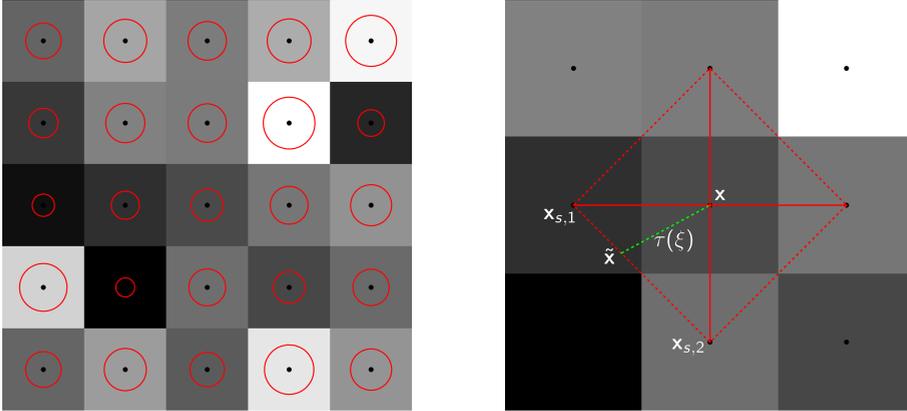


Figure 2.2: Tsitsiklis' 4-point stencil works for isotropic metrics.

for some $\mathcal{C} : \mathbb{R}^2 \rightarrow \mathbb{R}^+$, i.e., the cost of traversing $\mathbf{x} \in \mathbb{R}^2$ is independent of the orientation.

Let us say that in 2D problems, a stencil S in the point \mathbf{x} consists of N points, that form triangles as follows:

$$S(\mathbf{x}) := \{\mathbf{x}_1, \dots, \mathbf{x}_N\}, \quad (\Delta S)(\mathbf{x}) := \{\Delta \mathbf{x}_1 \mathbf{x}_2 \mathbf{x}, \Delta \mathbf{x}_2 \mathbf{x}_3 \mathbf{x}, \dots, \Delta \mathbf{x}_N \mathbf{x}_1 \mathbf{x}\}.$$

Tsitsiklis's method makes use of a four-point stencil on a Cartesian grid, see Fig. 2.2. It approximates Bellman's optimality principle

$$U(\mathbf{p}) = \inf_{\mathbf{q} \in \partial N} U(\mathbf{q}) + d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}),$$

with ∂N the boundary of a neighborhood N of \mathbf{q} .

The update procedure is as follows. Suppose we want to (re)compute the value of U at the point \mathbf{x} . Then with $\tilde{\mathbf{x}} = \xi \mathbf{x}_{s,1} + (1 - \xi) \mathbf{x}_{s,2}$, $\tau(\xi) = \|\tilde{\mathbf{x}} - \mathbf{x}\|$, we can write

$$U(\mathbf{x}) = \min_{s \in (\Delta S)(\mathbf{x})} V_s(\mathbf{x}),$$

$$V_s(\mathbf{x}) := \min_{\xi \in [0,1]} \{\mathcal{C}(\mathbf{x})\tau(\xi) + \xi U(\mathbf{x}_{s,1}) + (1 - \xi)U(\mathbf{x}_{s,2})\}.$$

In this case $S(\mathbf{x})$ consists of the four simplices (triangles) as in Fig. 2.2. This method and methods with a similar update strategy are called semi-Lagrangian, since they require the approximation of the value of U in points that are not part of the mesh.

It can be shown that for this metric and with this stencil, the distance map U can be consistently computed using the single-pass Fast Marching method, i.e., U would converge to the viscosity solution if the grid size converges to 0. A necessary condition for such consistency is *causality*.

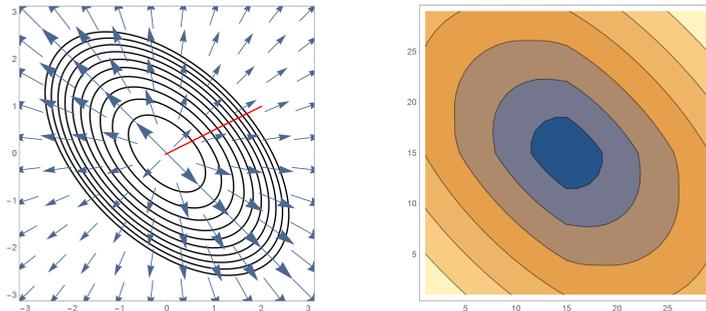


Figure 2.3: Ellipsoidal level set lines for the example mentioned above, with on the left a vector field corresponding to $M^{-1}\nabla U(\mathbf{x})$, for $c_1 = c_2 = 1$. On the right the solution as found with Tsitsiklis' algorithm.

Definition 2.3.2 (Causality). *A stencil is called causal in the point $\mathbf{p} \in \mathbb{M}$ if, for all points \mathbf{q} in the stencil that are used to compute $U(\mathbf{p})$ (i.e., all points that 'cause' the value $U(\mathbf{p})$), $U(\mathbf{p}) \geq U(\mathbf{q})$.*

However, the causality condition of the stencil might be violated when using anisotropic metrics. A nice and simple example to illustrate that this method is not suitable for anisotropic metrics is given in [SV01], where they consider the distance on a tilted plane:

Example. Let $U(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$ be the distance to the origin on the plane $z = c_1x + c_2y$, for some vector $\mathbf{c} = (c_1, c_2)$. We can define the optimization problem using the metric

$$\mathcal{F}(\mathbf{x}, \dot{\mathbf{x}}) = \sqrt{\dot{\mathbf{x}}, M\dot{\mathbf{x}}} =: \|\dot{\mathbf{x}}\|_M, \quad M^{-1} = \frac{1}{1 + c_1^2 + c_2^2} \begin{pmatrix} 1 + c_1^2 & -c_1c_2 \\ -c_1c_2 & 1 + c_2^2 \end{pmatrix}. \quad (2.30)$$

We get the HJB equation

$$\mathcal{F}^*(\mathbf{x}, dU(\mathbf{x})) = \sqrt{(\nabla U(\mathbf{x}), M^{-1}\nabla U(\mathbf{x}))} = 1, \quad (2.31)$$

(where $dU(\mathbf{x}) \in T_{\mathbf{x}}^*(\mathbb{M})$, whose Riesz representative (see e.g. [Rud06]) w.r.t the standard inner product equals $\nabla U(\mathbf{x}) \in T_{\mathbf{x}}(\mathbb{M})$). The exact solution to the problem is easy, as it is also just the Euclidean distance in \mathbb{R}^3 :

$$U(x, y) = \sqrt{x^2 + y^2 + (c_1x + c_2y)^2}. \quad (2.32)$$

It can be checked formally that this is indeed the viscosity solution the HJ equation. However, as can be seen in Fig. 2.3, the numerical solution shows some errors, that would not vanish even when the grid size is decreased.

For such semi-Lagrangian approaches, there is another definition that is directly related to the geometry of the stencils, that can be shown to be equivalent to the causality property.

Definition 2.3.3 ((Generalized) acuteness). *A stencil at point \mathbf{p} is acute when for all points \mathbf{q}, \mathbf{q}' in a common simplex of the stencil, the following holds:*

$$(\mathbf{q} - \mathbf{p}, d_{\mathbf{p}}\mathcal{F}(\mathbf{p}, \mathbf{q}' - \mathbf{p})) \geq 0,$$

where $d_{\mathbf{p}}$ denotes differentiation w.r.t. the second entry of \mathcal{F} .

For an isotropic metric as in (2.29), this definition reduces to the standard notion of acuteness $(\mathbf{q} - \mathbf{p}, \mathbf{q}' - \mathbf{p}) \geq 0$, which is indeed satisfied by Tsitsiklis's four-point stencil.

Adapting the stencil such that the causality property is recovered can be done with either structured or unstructured meshes. A method with unstructured methods was presented in [SV01]. However, since our goal is to use images as data, which are already defined on structured meshes, we prefer to use structured meshes throughout. A lot of work regarding the design of small stencils (in both number of points and stencil radius) for anisotropic metrics on structured meshes was done in [Mir13, Mir14].

As an alternative to the semi-Lagrangian approaches described above, there exist Eulerian Fast Marching approaches, that only require computation of values on the grid points. They discretize the Hamilton-Jacobi equation using finite differences, that determines how to update values of the distance function. In this case, the finite difference approximation should be monotone to approximate the unique viscosity solution instead of another weak solution. It should again be causal to ensure that it can be solved using the Fast Marching algorithm.

For the isotropic metric (2.29), the alternative to the four-point semi-Lagrangian approach can be formulated using only upwind differences. When we define U_{ij} to be the value of U at grid point (x_i, y_j) , and \mathcal{C}_{ij} the value of the external cost at that grid point, an upwind finite difference approximation of the Hamilton-Jacobi equation is:

$$\max(U_{ij} - U_{(i-1)j}, -(U_{(i+1)j} - U_{ij}), 0)^2 + \max(U_{ij} - U_{i(j-1)}, -(U_{i(j+1)} - U_{ij}), 0)^2 = h^2 \mathcal{C}_{ij}^2.$$

Note how this scheme is non-decreasing with respect to U_{ij} , and each of the differences, which makes the scheme monotone. The fact that it only depends on the positive parts of each of the differences makes the scheme causal: when updating, only values upwind of the wavefront are used to compute values downwind of the wavefront. This Eulerian scheme was shown to be closely connected to the semi-Lagrangian scheme [SV01].

Beyond this 2D case, in general the discretized version of the Hamilton-Jacobi equation is roughly of the form

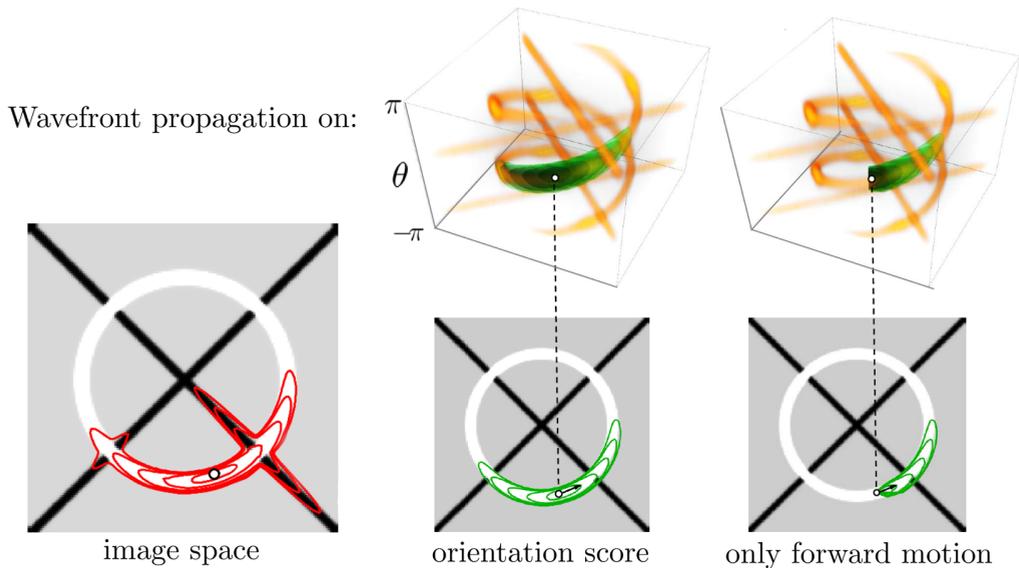


Figure 2.4: An example of (anisotropic) wavefront propagation for a 2D image. Left: propagation of wavefronts (red) with a metric based on the 2D image. Middle: propagation of wavefronts (green) on position-orientation space based on the orientation score of the image. Right: same as the middle, but allowing only forward propagation. The point (left) and point-and-arrow (middle, right) indicate the source points. The orange color in the orientation scores (top) indicate presence of structure, where the front can propagate faster.

$$\mathcal{F}^*(\mathbf{p}, dU(\mathbf{p})) \approx \text{FD}(\mathbf{p}, U(\mathbf{p}), (U(\mathbf{p}) - U(\mathbf{q}))_{\mathbf{q} \in S(\mathbf{p})}) = 1, \quad (2.33)$$

where FD is some function approximating the dual of \mathcal{F} based on finite differences. In this context, the scheme is *monotone* if FD is a non-decreasing function of $U(\mathbf{p})$ and $U(\mathbf{q})$ for $\mathbf{q} \in S(\mathbf{p})$, for all \mathbf{p} in the domain. The scheme is *causal* if FD only depends on the positive part of the differences $U(\mathbf{p}) - U(\mathbf{q})$ for all $\mathbf{q} \in S(\mathbf{p})$, for all \mathbf{p} . The design of such finite difference schemes in this Eulerian setting, for a range of Hamiltonians, was worked out in the recent work [Mir17a, Mir17b].

2.3.4 Motivation for wavefront propagation on positions and orientations

To illustrate the approach presented in this section, we conclude with an example. Consider the image in Fig. 2.4, consisting of a circle and a cross. The point in the image is the source point, say $\mathbf{x}_S \in \mathbb{R}^2$, where $U(\mathbf{x}_S) = 0$. Using an anisotropic metric $\mathcal{F}(\mathbf{x}, \dot{\mathbf{x}}) = \mathcal{C}(\mathbf{x}) \|\dot{\mathbf{x}}\|_{M(\mathbf{x})}$, with \mathcal{C} a cost function induced by the image and $M(\mathbf{x})$ a sym-

metric, positive definite matrix with eigenvectors along and perpendicular to the lines in the image. Such an M usually does not have a preferred direction on positions in the image where structures cross, resulting in an isotropic metric at those position.

As a result, the wavefronts that correctly follow the circular structure away from the source point, leak into the crossing structure. This eventually leads to problems in the segmentation of the image.

As was briefly discussed in Section 1.3, an approach to generically deal with such crossing structures is by constructing an orientation score from the image, i.e. lifting it to position-orientation space using an invertible transform [Dui05]. Structures with a different orientation are now separated to different orientation layers, see Fig. 2.4, where the orange color indicates structures on that position/orientation.

We can define an appropriate Finsler metric on the tangent space of the extended domain, $\mathcal{F} : T(\mathbb{R}^2 \times S^1) \rightarrow \mathbb{R}^+$. This metric should put a high cost on vertical motion through the orientation score (which corresponds to a high curvature in image space). The source point now has a position and orientation. By projecting the wavefronts (computed in the extended domain) back onto the image, we obtain the green wavefronts as in the middle of Fig. 2.4. This approach prevents the wavefronts from leaking from the circle into the cross.

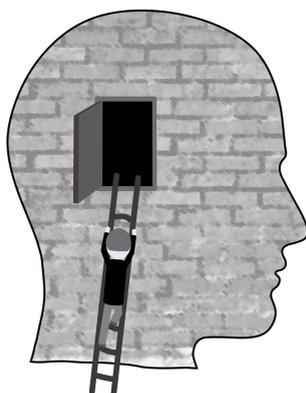
The above example emphasizes the need for the extension of this Hamilton-Jacobi framework to position-orientation space. The benefits of this approach have already been shown for 2D images in [BDMS15, SBD⁺15]. The extension to 3D images and dMRI images will be the topic of Chapters 5 and 7. We will further specify the type of metric that we use to obtain curvature penalized optimal paths in position-orientation space, with a cost function that is driven by the data. In those chapters it will also become important to allow only forward propagation of the wavefront, as illustrated on the right in Fig. 2.4.

Part II

Theoretical contributions

Solutions of (Convection-)Diffusion Kernels on $SE(3)$

Based on: [PD17] J. M. Portegies and R. Duits, “New Exact and Numerical Solutions of the (Convection-)Diffusion Kernels on $SE(3)$,” *Differential Geometry and its Applications*, vol. 53, pp. 182–219, Aug. 2017.



3.1 Introduction

The properties of the kernels of the hypo-elliptic (convection-)diffusions on Lie groups, in particular the Euclidean motion group, have been of interest in fields such as image analysis [Aug01, AZ03, CS06, vA07, MSS09, DF10b], robotics [Chi11] and harmonic analysis [tER98, CK00, ABGR09, SD15]. In [Mum94] Mumford posed the problem of finding solutions of the kernels for the convection-diffusion process (direction process) on the roto-translation group $SE(2)$. Subsequently, in [TW97] analytic approximations were provided. Several numerical approaches were provided in [Aug01, ZW04, San11, ZDStHR16, CK00] and exact solutions were derived in [DvA08].

Hypo-elliptic differential operators on Lie groups have been studied in great detail by [RS76], and existence of a kernel, along with estimates of the fundamental solution, was proven in [JS86]. In [ABGR09], a Fourier-based method is presented to compute the kernel of the hypo-elliptic heat equation on unimodular Lie groups and examples are provided for various Lie groups. Three approaches to derive the exact solutions for the heat kernel for $SE(2)$ have been proposed in $SE(2)$ in [DF09, DF10b], of which the first is equivalent to the approach in [ABGR09] and provides a solution in terms of a Fourier series. This approach will be extended in this chapter to $SE(3)$ and the connection to the Fourier transform on $SE(3)$ will be made. The second approach of [DF09, DF10b] provides a solution in terms of rapidly decaying functions and in the third approach this series is explicitly computed in terms of Mathieu functions.

The use of these diffusion and convection-diffusion equations for 2D images has already briefly been described in Section 1.2. How they can be used for enhancement of dMRI data will be further explored in Chapter 6. Furthermore, in [MSS09] the 3D extension of Mumford's direction process was used to enhance dMRI data with the aid of stochastic completion fields. In [RKK12] it is shown that the convection-diffusion kernel can be used to obtain asymmetric, regularized FODs. The practical advantages of the diffusion process for DTI, in particular better fiber tractography results and improved connectivity measures, are given in [PRD⁺10, PAV⁺15].

Although finite-difference implementations of the PDEs exist [CDDH11], as well as kernel approximations [RS76, JS86, DF10a] (see also Chapter 4), so far no exact expressions are known. The derivation of these exact solutions will be one of the main results of this chapter. Other contributions of this chapter are summarized at the end of this introduction. We first recall some definitions and notation.

3.1.1 Legal (convection-)diffusion operators on $\mathbb{R}^3 \times S^2$

In Section 2.2.2 we have discussed the use of legal operators on $\mathbb{R}^3 \times S^2$, and thereby reduced our equations of interest to

$$\begin{cases} \partial_t W(\mathbf{y}, \mathbf{n}, t) = (Q_i W)(\mathbf{y}, \mathbf{n}, t), \\ W(\mathbf{y}, \mathbf{n}, 0) = U(\mathbf{y}, \mathbf{n}), \end{cases} \quad (3.1)$$

with

$$Q_1 := D_{33}(\mathbf{n} \cdot \nabla_{\mathbb{R}^3})^2 + D_{44}\Delta_{S^2} \quad (3.2)$$

for pure diffusion and

$$Q_2 := -(\mathbf{n} \cdot \nabla_{\mathbb{R}^3}) + D_{44}\Delta_{S^2}. \quad (3.3)$$

for convection-diffusion. Here $\nabla_{\mathbb{R}^3}$ denotes the gradient on \mathbb{R}^3 , and Δ_{S^2} is the Laplace-Beltrami operator on the sphere S^2 . As before, $(\mathbf{y}, \mathbf{n}) \in \mathbb{R}^3 \times S^2$, $t \geq 0$. In particular, we want to find an (at least formally) exact solution for the impulse response or convolution kernel of these PDEs, i.e., the solution with initial condition $U(\mathbf{y}, \mathbf{n}) = \delta_{(\mathbf{0}, \mathbf{e}_z)}(\mathbf{y}, \mathbf{n})$. We denote the linear and bounded evolution operator that maps $U(\mathbf{y}, \mathbf{n})$ on $W(\mathbf{y}, \mathbf{n}, t)$ by $\Upsilon_t : \mathbb{L}_2(\mathbb{R}^3 \times S^2) \rightarrow \mathbb{L}_2(\mathbb{R}^3 \times S^2)$:

$$W(\mathbf{y}, \mathbf{n}, t) = (\Upsilon_t U)(\mathbf{y}, \mathbf{n}), \quad \text{for all } (\mathbf{y}, \mathbf{n}) \in \mathbb{R}^3 \times S^2. \quad (3.4)$$

As mentioned in Section 2.2.3, the kernel for a certain evolution time of the PDE is also the probability density function of a random walker starting in the origin that travels the same amount of time. When we are interested in the position of a random walker regardless of the traveling time, we can condition on the traveling time and integrate. If we assume to have exponentially distributed traveling time $t \sim \text{Exp}(\alpha)$, with mean $1/\alpha$, $\alpha > 0$, then the probability density function of a random particle given its initial distribution U can be written as:

$$P_\alpha(\mathbf{y}, \mathbf{n}) = \alpha \int_0^\infty e^{-t\alpha} (e^{t(Q^{\mathbf{D}, \mathbf{a}})} U)(\mathbf{y}, \mathbf{n}) dt = -\alpha(Q^{\mathbf{D}, \mathbf{a}} - \alpha I)^{-1} U(\mathbf{y}, \mathbf{n}). \quad (3.5)$$

This shows that the time-integrated process relates to the resolvent operator of $Q^{\mathbf{D}, \mathbf{a}}$, in particular for Q_1 and Q_2 in Eq. (3.2) and (3.3), respectively. Therefore, apart from the time evolutions, we are also concerned with the corresponding resolvent equation:

$$((Q_i - \alpha I) P_\alpha^i)(\mathbf{y}, \mathbf{n}) = -\alpha U(\mathbf{y}, \mathbf{n}), \quad (3.6)$$

with $(\mathbf{y}, \mathbf{n}) \in \mathbb{R}^3 \times S^2$, $\alpha > 0$, $i = 1, 2$.

Spectral decomposition of generator Q_i (with a purely discrete spectrum) therefore also yields the spectral decomposition of both the (compact) operator Υ_t and the

(compact) resolvent operator $(Q_i - \alpha I)^{-1}$. Furthermore, it can be shown that the resolvent operator occurs in Tikhonov regularization of the input [FDB04, DF10a]. For the convection-diffusion case, the time-integrated process is practically even more useful than the time-dependent case.

3.1.2 Contributions and outline

In this chapter, we provide exact expressions for the kernels corresponding to the generators Q_1 and Q_2 for diffusion and convection-diffusion, in terms of eigenfunctions of the operator, after applying a Fourier transform in the spatial coordinates. As such we solve the 3D version of Mumford's direction process [Mum94], and the 3D version of the (2D) hypo-elliptic Brownian motion kernel, considered for orientation processing in image analysis in [CS06, ABGR09, DF10b]. The main results can be found in Theorems 3.2.3, 3.3.3 and 3.3.4 and in Cor. 3.2.5.

A similar approach was used in [DvA08] for the $SE(2)$ -case, but for the $SE(3)$ -case exact solutions are not known, to the best of our knowledge. For $SE(2)$, solutions were expressed in [DvA08] in terms of Mathieu functions. In $SE(3)$, we encounter (generalized) spheroidal wave functions, that can be written as a series of associated Legendre functions. We use the eigenfunctions to give expressions for both the time-dependent and the time-integrated process, associated with a resolvent equation. Using the results for the exact solutions for the time-integrated case, we derive in Thm 3.3.6 the minimal amount of repeated convolutions of the resolvent kernels required to remove the singularities in the origin. The proof of this result is given in Appendix A.3.

We also provide a numerical method to approximate the kernels, that is an extension of the numerical algorithm by August [Aug01] from $SE(2)$ to $SE(3)$. This algorithm has the advantage that it is closely related to the exact solutions and can therefore be shown to yield convergence to the exact solutions. A full comparison between exact solutions, stochastic process limits and kernel approximations for these processes on $SE(2)$ has been done in [ZDStHR16]. In this chapter, the focus will be on the derivation of the exact solutions and its connection to a numerical approximation method that generalizes the results [Aug01, DvA08] from $SE(2)$ to $SE(3)$.

Finally, in order to make the connection to earlier work on exact solutions of heat kernels in the $SE(2)$ -case [ABGR09, DvA08] via harmonic analysis and the Fourier transform on $SE(2)$, we also derive equivalent representations of the kernels using the $SE(3)$ Fourier transform. The details of this equivalent representation (and alternative roadmap to the solutions) is presented in Appendix A.4, where we follow an algebraic Fourier theoretic approach. We stress that the body of this chapter relies on classical, geometrical and functional analysis. However, the approach in Appendix A.4 provides the reader further insight in specific choices in the geometric analysis in this chapter.

The chapter is structured as follows. In Sections 3.2 and 3.3 we derive exact expressions for the convolution kernels of the differential equations in (3.1), with the main results in Theorem 3.2.3 and Theorem 3.3.4. We emphasize that the roadmap and computations in these sections are very similar, only in the convergence proofs of the encountered series of eigenfunctions we need different theory for each case. In Section 3.4 we present a matrix representation of the evolution in a Fourier basis and provide an algorithm to numerically compute a truncation of the exact series solution. We summarize our findings and conclude in Section 3.5. The derivation of equivalent solutions for the main PDEs using the SE(3) Fourier transform can be found in Appendix A.4.

3.2 Derivation of the exact solutions for hypo-elliptic diffusion on $\mathbb{R}^3 \times S^2$

In this section we derive the exact solutions for the hypo-elliptic diffusion case. We first set out the formal procedure for finding these solutions, before we present the details for this particular case and the specific eigenfunctions that we encounter. The evolution process for hypo-elliptic diffusion on $\mathbb{R}^3 \times S^2$, i.e., with generator Q_1 as in (3.2), is written as follows:

$$\begin{cases} \partial_t W(\mathbf{y}, \mathbf{n}, t) = Q_1 W(\mathbf{y}, \mathbf{n}, t) = (D_{33}(\mathbf{n} \cdot \nabla_{\mathbb{R}^3})^2 + D_{44} \Delta_{S^2}) W(\mathbf{y}, \mathbf{n}, t), & t \geq 0 \\ W(\mathbf{y}, \mathbf{n}, 0) = U(\mathbf{y}, \mathbf{n}), \end{cases} \quad (3.7)$$

with $\mathbf{y} \in \mathbb{R}^3$, $\mathbf{n} \in S^2$. As in (3.2),(3.3), we use $\nabla_{\mathbb{R}^3}$ to indicate the gradient with respect to the spatial variables, and Δ_{S^2} to denote the Laplace-Beltrami operator on the sphere. Parameters $D_{33}, D_{44} > 0$ influence the amount of spatial and angular regularization, respectively. We use both a subscript and superscript 1 throughout this section for operators that arise from this evolution, to distinguish these from the operators corresponding to the convection-diffusion that we encounter in Section 3.3.

The domain of the generator Q_1 equals

$$\mathcal{D}(Q_1) = \mathbb{H}_2(\mathbb{R}^3) \otimes \mathbb{H}_2(S^2), \quad (3.8)$$

where we use \mathbb{H}_2 to denote a Sobolev space, although in $\mathcal{D}(Q_1)$ both $\mathbb{H}_2(\mathbb{R}^3)$ and $\mathbb{H}_2(S^2)$ are equipped with the usual \mathbb{L}_2 -norm. By linearity and left-invariance of the differential equation, the solution can be found by $\mathbb{R}^3 \times S^2$ -convolution with the corresponding integrable kernel $K_t^1 : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}^+$:

$$\begin{aligned}
W(\mathbf{y}, \mathbf{n}, t) &= (K_t^1 *_{\mathbb{R}^3 \times S^2} U)(\mathbf{y}, \mathbf{n}) \\
&= \int_{S^2} \int_{\mathbb{R}^3} K_t^1(\mathbf{R}_{\mathbf{n}'}^T(\mathbf{y} - \mathbf{y}'), \mathbf{R}_{\mathbf{n}'}^T \mathbf{n}) U(\mathbf{y}', \mathbf{n}') \, d\mathbf{y}' \, d\sigma(\mathbf{n}'). \tag{3.9}
\end{aligned}$$

The specific choice for the rotation matrix $\mathbf{R}_{\mathbf{n}'}$ does not matter, since the left-invariance of the PDE implies that $K_t^1(\mathbf{y}, \mathbf{n}) = K_t^1(\mathbf{R}_{\mathbf{e}_z, \alpha} \mathbf{y}, \mathbf{R}_{\mathbf{e}_z, \alpha} \mathbf{n})$ for all $\alpha \in [0, 2\pi]$, see [DF10a] and Section 2.1.1. Our approach for finding the exact kernel K_t^1 is inspired by the approach for the $SE(2)$ case in [DvA08]. We first apply a Fourier transform with respect to the spatial variables:

$$\hat{W}(\boldsymbol{\omega}, \mathbf{n}, t) := \mathcal{F}_{\mathbb{R}^3}(W)(\boldsymbol{\omega}, \mathbf{n}, t) = \int_{\mathbb{R}^3} W(\mathbf{y}, \mathbf{n}, t) e^{-i\boldsymbol{\omega} \cdot \mathbf{y}} \, d\mathbf{y}. \tag{3.10}$$

The hat is used to indicate that a function has been Fourier transformed. The PDE (3.7) in terms of \hat{W} then becomes:

$$\begin{cases} \partial_t \hat{W}(\boldsymbol{\omega}, \mathbf{n}, t) = (D_{44} \Delta_{S^2} - D_{33}(\boldsymbol{\omega} \cdot \mathbf{n})^2) \hat{W}(\boldsymbol{\omega}, \mathbf{n}, t), \\ \hat{W}(\boldsymbol{\omega}, \mathbf{n}, 0) = \mathcal{F}_{\mathbb{R}^3}(U)(\boldsymbol{\omega}, \mathbf{n}) =: \hat{U}(\boldsymbol{\omega}, \mathbf{n}). \end{cases} \tag{3.11}$$

We fix $\boldsymbol{\omega} \in \mathbb{R}^3$ and we define the operator $\mathcal{B}_{\boldsymbol{\omega}}^1 : \mathbb{H}_2(S^2) \rightarrow \mathbb{L}_2(S^2)$ as follows:

$$\mathcal{B}_{\boldsymbol{\omega}}^1 := D_{44} \Delta_{S^2} - D_{33}(\boldsymbol{\omega} \cdot \mathbf{n})^2. \tag{3.12}$$

We use the subscript $\boldsymbol{\omega}$ to explicitly indicate that the operator depends on the frequency vector $\boldsymbol{\omega}$, as will the eigenfunctions of this operator. When we write

$$(\mathcal{B}\hat{W})(\boldsymbol{\omega}, \mathbf{n}) := (\mathcal{B}_{\boldsymbol{\omega}}^1 \hat{W}(\boldsymbol{\omega}, \cdot))(\mathbf{n}), \tag{3.13}$$

then the correspondence between the operator \mathcal{B} and the generator Q_1 can be written as:

$$\mathcal{B} = (\mathcal{F}_{\mathbb{R}^3} \otimes \mathbb{1}_{\mathbb{L}_2(S^2)}) \circ Q_1 \circ (\mathcal{F}_{\mathbb{R}^3}^{-1} \otimes \mathbb{1}_{\mathbb{H}_2(S^2)}). \tag{3.14}$$

The heat kernel $\hat{K}_t^1 \in \mathbb{L}_2(\mathbb{R}^3 \times S^2) \cap C(\mathbb{R}^3 \times S^2)$ of the PDE in the Fourier domain should then satisfy:

$$\begin{cases} \partial_t \hat{K}_t^1(\boldsymbol{\omega}, \mathbf{n}) = \mathcal{B}_{\boldsymbol{\omega}}^1 \hat{K}_t^1(\boldsymbol{\omega}, \mathbf{n}), \\ \hat{K}_0^1(\boldsymbol{\omega}, \mathbf{n}) = \delta_{\mathbf{e}_z}(\mathbf{n}), \end{cases} \tag{3.15}$$

with $\delta_{\mathbf{e}_z}$ the δ -distribution on S^2 at \mathbf{e}_z . We show later that the $\mathbb{L}_2(S^2)$ -normalized eigenfunctions of the operator $\mathcal{B}_{\boldsymbol{\omega}}^1$ form an orthonormal basis for $\mathbb{L}_2(S^2)$ and that, similar

to the enumeration of spherical harmonics, these functions are indexed with integers l and m , $|m| \leq l$. For the eigenfunctions, that we denote with $\Phi_{l,m}^\omega$, we have:

$$\mathcal{B}_\omega^1 \Phi_{l,m}^\omega = \lambda_\omega^{l,m} \Phi_{l,m}^\omega, \quad \text{with } \lambda_\omega^{l,m} \leq 0. \quad (3.16)$$

The kernel \hat{K}_t^1 can then be written in terms of these eigenfunctions as

$$\hat{K}_t^1(\omega, \mathbf{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \overline{\Phi_{l,m}^\omega(\mathbf{e}_z)} \Phi_{l,m}^\omega(\mathbf{n}) e^{\lambda_\omega^{l,m} t}, \quad (3.17)$$

where the overline indicates complex conjugation. The solution of the differential equation (3.11) in the Fourier domain is given by:

$$\hat{W}(\omega, \mathbf{n}, t) = \sum_{l=0}^{\infty} \sum_{m=-l}^l (\hat{U}(\omega, \cdot), \Phi_{l,m}^\omega)_{\mathbb{L}_2(S^2)} \Phi_{l,m}^\omega(\mathbf{n}) e^{\lambda_\omega^{l,m} t}, \quad (3.18)$$

where we rely on the following inner product convention on $\mathbb{L}_2(S^2)$:

$$(f, g)_{\mathbb{L}_2(S^2)} = \int_{S^2} f(\mathbf{n}) \overline{g(\mathbf{n})} d\sigma(\mathbf{n}). \quad (3.19)$$

Thereby, the solution of Eq. (3.7) is given by:

$$W(\mathbf{y}, \mathbf{n}, t) = \left[\mathcal{F}_{\mathbb{R}^3}^{-1} \hat{W}(\cdot, \mathbf{n}, t) \right] (\mathbf{y}). \quad (3.20)$$

This expression should coincide with the convolution expression in Eq. (3.9). The following lemma gives us a useful identity for the eigenfunctions $\Phi_{l,m}^\omega$, that allows us to connect the series expression for the solution of (3.7) with the convolution form in (3.9).

Lemma 3.2.1. *For all $l \in \mathbb{N}_0, m \in \mathbb{Z}, |m| \leq l$, let $\Phi_{l,m}^\omega(\mathbf{n})$ be an eigenfunction of \mathcal{B}_ω^1 , with eigenvalue $\lambda_\omega^{l,m}$, and let $\mathbf{R} \in SO(3)$. Then $\Phi_{l,m}^\omega(\mathbf{R}^T \cdot)$ is an eigenfunction of $\mathcal{B}_{\mathbf{R}\omega}^1$ with eigenvalue $\lambda_{\mathbf{R}\omega}^{l,m} = \lambda_\omega^{l,m}$.*

Proof. We can write

$$\begin{aligned} (\mathcal{B}_{\mathbf{R}\omega}^1 \Phi_{l,m}^\omega(\mathbf{R}^T \cdot))(\mathbf{n}) &= (\Delta_{S^2} \Phi_{l,m}^\omega(\mathbf{R}^T \cdot))(\mathbf{n}) - (\mathbf{R}\omega, \mathbf{n})^2 (\Phi_{l,m}^\omega)(\mathbf{R}^T \mathbf{n}) \\ &= (\Delta_{S^2} \Phi_{l,m}^\omega)(\mathbf{R}^T \mathbf{n}) - (\omega, \mathbf{R}^T \mathbf{n})^2 \Phi_{l,m}^\omega(\mathbf{R}^T \mathbf{n}) \\ &= (\mathcal{B}_\omega^1 \Phi_{l,m}^\omega(\cdot))(\mathbf{R}^T \mathbf{n}) \\ &= \lambda_\omega^{l,m} \Phi_{l,m}^\omega(\mathbf{R}^T \mathbf{n}), \end{aligned} \quad (3.21)$$

from which the result follows. \square

From the lemma we conclude that the eigenvalues $\lambda_{\boldsymbol{\omega}}^{l,m}$ only depend on the norm $r = \|\boldsymbol{\omega}\|$ of the frequency $\boldsymbol{\omega}$, so from now on we write $\lambda_r^{l,m}$. Moreover, we have the following relation between eigenfunctions:

$$\Phi_{l,m}^{\mathbf{R}\boldsymbol{\omega}}(\mathbf{n}) = \Phi_{l,m}^{\boldsymbol{\omega}}(\mathbf{R}^T \mathbf{n}). \quad (3.22)$$

We can combine the previous considerations and Lemma 3.2.1 to give two equivalent series expressions for the kernel \hat{K}_1 satisfying Eq. (3.15):

$$\hat{K}_t^{-1}(\boldsymbol{\omega}, \mathbf{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \overline{\Phi_{l,m}^{\boldsymbol{\omega}}(\mathbf{e}_z)} \Phi_{l,m}^{\boldsymbol{\omega}}(\mathbf{n}) e^{\lambda_r^{l,m} t} = \sum_{l=0}^{\infty} \sum_{m=-l}^l \overline{\Phi_{l,m}^{\mathbf{R}_{\mathbf{n}'}\boldsymbol{\omega}}(\mathbf{n}')} \Phi_{l,m}^{\mathbf{R}_{\mathbf{n}'}\boldsymbol{\omega}}(\mathbf{R}_{\mathbf{n}'} \mathbf{n}) e^{\lambda_r^{l,m} t}. \quad (3.23)$$

Finally, the solution of (3.7) can then be written as

$$\begin{aligned} W(\mathbf{y}, \mathbf{n}, t) &= (K_t^1 *_{\mathbb{R}^3 \rtimes S^2} U)(\mathbf{y}, \mathbf{n}) = ((\mathcal{F}_{\mathbb{R}^3}^{-1} \hat{K}_t^{-1}(\cdot, \cdot)) *_{\mathbb{R}^3 \rtimes S^2} U)(\mathbf{y}, \mathbf{n}) \\ &= \int_{\mathbb{R}^3} \sum_{l=0}^{\infty} \sum_{m=-l}^l (\hat{U}(\boldsymbol{\omega}, \cdot), \Phi_{l,m}^{\boldsymbol{\omega}})_{\mathbb{L}_2(S^2)} \Phi_{l,m}^{\boldsymbol{\omega}}(\mathbf{n}) e^{\lambda_r^{l,m} t} e^{i\mathbf{y} \cdot \boldsymbol{\omega}} d\boldsymbol{\omega}. \end{aligned} \quad (3.24)$$

Now that we have formal expressions for the solution of the hypo-elliptic diffusion PDE, we can focus on finding analytic expressions for the eigenfunctions $\Phi_{l,m}^{\boldsymbol{\omega}}$.

3.2.1 Frequency-dependent choice of spherical variables

So far we have not specified a choice of spherical variables for the functions to which the operator $\mathcal{B}_{\boldsymbol{\omega}}^1$ is applied. This is needed in order to derive expressions for the eigenfunctions $\Phi_{l,m}^{\boldsymbol{\omega}}$. As in the case for spherical harmonics, we want to use separation of variables for each fixed spatial frequency $\boldsymbol{\omega}$. To be able to do this, we choose to parameterize the orientation of \mathbf{n} using angles dependent on $\boldsymbol{\omega}$. This choice should be such that the variables can be separated in both the Laplace-Beltrami operator Δ_{S^2} and in the multiplication operator $(\boldsymbol{\omega} \cdot \mathbf{n})^2$. Note that the standard spherical coordinates are not suitable for this purpose.

To this end, we choose spherical coordinates with respect to the (normalized) frequency $r^{-1}\boldsymbol{\omega}$, with $r = \|\boldsymbol{\omega}\|$, and a second axis perpendicular to $\boldsymbol{\omega}$. The specific choice for the latter axis is not important, since in our PDE only the angle between $\boldsymbol{\omega}$ and \mathbf{n} plays a role. For convenience we take as second axis $\frac{\boldsymbol{\omega} \times \mathbf{e}_z}{\|\boldsymbol{\omega} \times \mathbf{e}_z\|}$, and we let β, γ denote the angles of rotation about axes $\frac{\boldsymbol{\omega} \times \mathbf{e}_z}{\|\boldsymbol{\omega} \times \mathbf{e}_z\|}$ and $r^{-1}\boldsymbol{\omega}$, respectively. For $r^{-1}\boldsymbol{\omega} = \mathbf{e}_z$, β and γ are just the standard spherical coordinates. Every orientation $\mathbf{n} \in S^2$ can now be written in the form

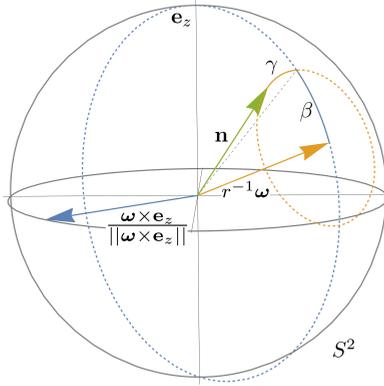


Figure 3.1: For $\omega \neq e_z$, we parameterize every orientation \mathbf{n} (green) by rotations around $r^{-1}\omega$ (orange) and $\frac{\omega \times e_z}{\|\omega \times e_z\|}$ (blue). In other words, $\mathbf{n}^\omega(\beta, \gamma) = \mathbf{R}_{r^{-1}\omega, \gamma} \mathbf{R}_{\frac{\omega \times e_z}{\|\omega \times e_z\|}, \beta}(r^{-1}\omega)$.

$$\mathbf{n} = \mathbf{n}^\omega(\beta, \gamma) = \mathbf{R}_{r^{-1}\omega, \gamma} \mathbf{R}_{\frac{\omega \times e_z}{\|\omega \times e_z\|}, \beta}(r^{-1}\omega), \quad \text{with } r = \|\omega\|, \quad (3.25)$$

see Fig. 3.1.

Lemma 3.2.2. *The Laplace-Beltrami operator on S^2 with the choice of variables as in Eq. (3.25) is given by:*

$$\Delta_{S^2} = \frac{1}{\sin \beta} \frac{\partial}{\partial \beta} \left(\sin \beta \frac{\partial}{\partial \beta} \right) + \frac{1}{\sin^2 \beta} \frac{\partial^2}{\partial \gamma^2}. \quad (3.26)$$

The full operator \mathcal{B}_ω^1 as defined in (3.12) with this choice of coordinates takes the separable form

$$\mathcal{B}_\omega^1 = \frac{D_{44}}{\sin \beta} \frac{\partial}{\partial \beta} \left(\sin \beta \frac{\partial}{\partial \beta} \right) + \frac{D_{44}}{\sin^2 \beta} \frac{\partial^2}{\partial \gamma^2} - D_{33}(r \cos \beta)^2. \quad (3.27)$$

Proof. The result follows from direct computation of the metric tensor \mathcal{G} w.r.t. the coordinates in (3.25). \square

3.2.2 Separation of variables

Thanks to the choice of coordinates (3.25) and Lemma 3.2.2, we can apply the method of separation of variables to solve the diffusion equation (3.11). We first look for solutions of the PDE without initial conditions, and we take the solutions of the separable form $T(t)\Phi(\beta, \gamma)$:

$$\partial_t(T(t)\Phi(\beta, \gamma)) = \mathcal{B}_\omega^1(T(t)\Phi)(\beta, \gamma) \iff \Phi(\beta, \gamma)\partial_t T(t) = T(t)(\mathcal{B}_\omega^1\Phi)(\beta, \gamma). \quad (3.28)$$

It follows that we get

$$\frac{1}{T(t)} \partial_t T(t) = \frac{(\mathcal{B}_\omega^1 \Phi)(\beta, \gamma)}{\Phi(\beta, \gamma)} =: \lambda_r, \quad (3.29)$$

where λ_r is the separation constant. We get that $T(t) = K e^{\lambda_r t}$, K constant and λ_r an eigenvalue of \mathcal{B}_ω^1 .

For finding eigenfunctions $\Phi(\beta, \gamma)$ of \mathcal{B}_ω^1 , we assume that these functions can be written as

$$\Phi(\beta, \gamma) = B(\beta) C(\gamma). \quad (3.30)$$

We then find:

$$\left\{ \frac{\sin \beta}{B(\beta)} \frac{d}{d\beta} \left(\sin \beta \frac{dB(\beta)}{d\beta} \right) - \left(\frac{D_{33}}{D_{44}} r^2 \cos^2 \beta + \frac{\lambda_r}{D_{44}} \right) \sin^2 \beta \right\} + \frac{1}{C(\gamma)} \frac{d^2 C(\gamma)}{d\gamma^2} = 0. \quad (3.31)$$

The resulting equation for $C(\gamma)$ can be solved straightforwardly. Taking into account the 2π -periodicity of C , we find that $C(\gamma)$ is a multiple of $e^{im\gamma}$ for some $m \in \mathbb{Z}$. Normalization then gives:

$$C(\gamma) \equiv C_m(\gamma) = \frac{1}{\sqrt{2\pi}} e^{im\gamma}, \quad m \in \mathbb{Z}. \quad (3.32)$$

3.2.3 Spheroidal wave equation

The equation for $B(\beta)$, with separation constant m as above, now becomes:

$$\sin \beta \frac{d}{d\beta} \left(\sin \beta \frac{dB(\beta)}{d\beta} \right) + \left[- \left(\frac{D_{33}}{D_{44}} r^2 \cos^2 \beta + \frac{\lambda_r}{D_{44}} \right) \sin^2 \beta - m^2 \right] B(\beta) = 0. \quad (3.33)$$

With the substitution $x = \cos \beta$, $y(x) = B(\beta)$ (which is commonly done for equations of this type), we get:

$$(1-x^2) \frac{d^2 y(x)}{dx^2} - 2x \frac{dy(x)}{dx} + \left[- \frac{D_{33}}{D_{44}} r^2 x^2 - \frac{\lambda_r}{D_{44}} - \frac{m^2}{1-x^2} \right] y(x) = 0, \quad -1 \leq x \leq 1. \quad (3.34)$$

We introduce two more parameters to bring this equation in a standard form:

$$\rho := \sqrt{\frac{D_{33}}{D_{44}}} r, \quad \tilde{\lambda}_\rho = - \frac{\lambda_r}{D_{44}}. \quad (3.35)$$

We then find:

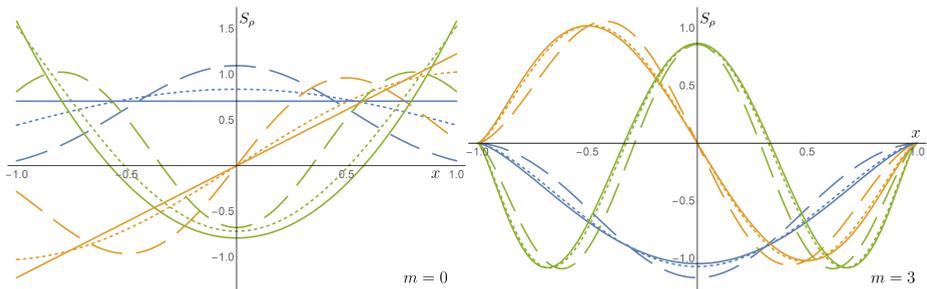


Figure 3.2: Plot of the spheroidal wave functions $S_\rho^{l,m}(x)$ for $m = 0$ (left) and $m = 3$ (right), with $\rho = 0, 2, 5$ (indicated with solid, small dashed and long dashed lines, respectively) and $l = m, m + 1, m + 2$ (indicated with blue, yellow and green, respectively).

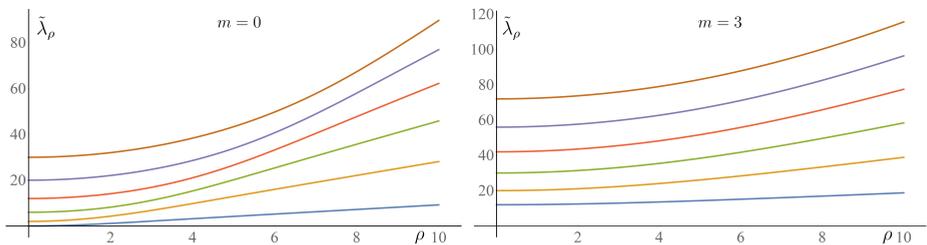


Figure 3.3: Plot of the spheroidal eigenvalues $\tilde{\lambda}_\rho^{l,m}$ as a function of ρ , for $m = 0$ (left) and $m = 3$ (right), and $l = m, \dots, m + 4$ (top to bottom). The eigenvalues are real for all ρ .

$$\frac{d}{dx} \left[(1 - x^2) \frac{dy(x)}{dx} \right] + \left[\tilde{\lambda}_\rho - \rho^2 x^2 - \frac{m^2}{1 - x^2} \right] y(x) = 0. \quad (3.36)$$

This equation is known as the spheroidal wave equation (SWE) [Fla57, OLBC10, Eq. 30.2.1], for which the eigenvalues and eigenfunctions, commonly referred to as spheroidal eigenvalues and spheroidal wave functions, are known and can be found up to arbitrary accuracy. In the remainder of this chapter, we denote them with $\tilde{\lambda}_\rho^{l,m}$ and $S_\rho^{l,m}$, respectively, for $l \in \mathbb{N}_0$, $m \in \mathbb{Z}$, $|m| \leq l$. For explicit analytic representations, see (A.11) in Appendix A.1.

The (normalized) spheroidal wave functions and the spheroidal eigenvalues as a result of our computations are displayed in Figs. 3.2 and 3.3 for a selection of values of l and m . It can be seen that all eigenvalues are real and all eigenfunctions are real-valued and vary continuously with parameter ρ . In the next section, we show that the spheroidal wave functions form a complete orthonormal basis for $\mathbb{L}_2([-1, 1])$, from which it follows that the eigenfunctions $\Phi_{l,m}^\omega$ form a complete orthonormal basis for $\mathbb{L}_2(S^2)$.

3.2.4 Sturm-Liouville form

The spheroidal wave equation can be written in Sturm-Liouville form:

$$(Ly)(x) = \frac{d}{dx} \left[p(x) \frac{dy(x)}{dx} \right] + q(x)y(x) = -\tilde{\lambda}_\rho w(x)y(x), \quad x \in [-1, 1]. \quad (3.37)$$

For this we choose $p(x) = (1 - x^2)$, $q(x) = -\rho^2 x^2 - \frac{m^2}{1-x^2}$, and we have weight function $w(x) = 1$. In this formulation, $p(x)$ vanishes at the boundary of the interval, which makes our problem a singular Sturm-Liouville problem (on a finite interval). It is sufficient to require boundedness of the solution and its derivative at the boundary points to have nonnegative, distinct, simple eigenvalues and existence of a countable, complete orthonormal basis of eigenfunctions $\{y_k\}_{k=1}^\infty$ [MM56] for the spheroidal wave equation. Since the weight function $w(x) = 1$, orthogonality is understood in the following sense:

$$\int_{-1}^1 y_k(x) \overline{y_l(x)} w(x) dx = \int_{-1}^1 y_k(x) \overline{y_l(x)} dx = \delta_{kl}. \quad (3.38)$$

For our choice of $p(x), q(x)$ and $w(x)$, with m fixed, we have $y_k(\cos \beta) = S_\rho^{l,m}(\cos \beta)$ (with $k = l^2 + l + 1 + m$) and corresponding eigenvalues $\tilde{\lambda}_\rho^{l,m}$.

3.2.5 Main theorem

From the above considerations, we can come to the main result of this section.

Theorem 3.2.3. *The normalized eigenfunctions $\Phi_{l,m}^\omega$ of the operator \mathcal{B}_ω^1 in (3.27) are given by:*

$$\Phi_{l,m}^\omega(\mathbf{n}^\omega(\beta, \gamma)) = S_\rho^{l,m}(\cos \beta) C_m(\gamma), \quad l \in \mathbb{N}_0, m \in \mathbb{Z}, |m| \leq l, \quad \rho = \sqrt{\frac{D_{33}}{D_{44}}} \|\omega\|. \quad (3.39)$$

Here $S_\rho^{l,m}$ is the $\mathbb{L}_2([-1, 1])$ -normalized eigenfunction of Eq. (3.36), given in (A.11), with corresponding eigenvalues $\lambda_r^{l,m} = -D_{44} \tilde{\lambda}_\rho^{l,m}$, with $\tilde{\lambda}_\rho^{l,m}$ the standard eigenvalues of the SWE as in Eq. (3.36). Function C_m is as in Eq. (3.32). The solution of the hypo-elliptic diffusion on $\mathbb{R}^3 \times S^2$ Eq. (3.7) is given by:

$$W(\mathbf{y}, \mathbf{n}, t) = (K_t^1 *_{\mathbb{R}^3 \times S^2} U)(\mathbf{y}, \mathbf{n}), \quad \text{with} \quad (3.40)$$

$$K_t^1(\mathbf{y}, \mathbf{n}) = \mathcal{F}_{\mathbb{R}^3}^{-1} \left(\omega \mapsto \sum_{l=0}^{\infty} \sum_{m=-l}^l \overline{\Phi_{l,m}^\omega(\mathbf{e}_z)} \Phi_{l,m}^\omega(\mathbf{n}) e^{\lambda_r^{l,m} t} \right) (\mathbf{y}).$$

Proof. From the separation of variables approach above, and Eq. (3.38), it follows that the $\Phi_{l,m}^\omega$ are indeed normalized eigenfunctions. From the fact that the resolvent of

operator B_{ω}^1 is compact, which follows from Sturm-Liouville theory, we obtain, using the spectral decomposition of compact self-adjoint operators, a complete orthonormal basis of eigenfunctions of $\mathbb{L}_2(\mathbb{R}^3 \times S^2)$. As $\dim(S^2) = 2$, we can number the orthonormal basis with two indices l, m , similar to the spherical harmonics, that appear in the special case $\omega = \mathbf{0}$. This allows us to use the expression in (3.17) and the result follows. \square

In the particular case of $\omega = 0$, the operator B_{ω}^1 reduces to $B_0^1 = D_{44}\Delta_{S^2}$, for which the spherical harmonic functions are the eigenfunctions. For the spherical harmonics, we use the following convention:

$$\Delta_{S^2} Y^{l,m}(\beta, \gamma) = -l(l+1)Y^{l,m}(\beta, \gamma), \quad Y^{l,m}(\beta, \gamma) = \frac{\varepsilon_m}{\sqrt{2\pi}} P_l^m(\cos \beta) e^{im\gamma}, \quad (3.41)$$

with the associated Legendre polynomials P_l^m as defined in (A.1), and

$$\varepsilon_m = \begin{cases} (-1)^m & m \geq 0, \\ 1 & m < 0. \end{cases} \quad (3.42)$$

In the following corollary we write the eigenfunctions $\Phi_{l,m}^{\omega}$ in (3.39) directly in terms of the spherical harmonics.

Corollary 3.2.4. *The $\mathbb{L}_2(\mathbb{R}^3 \times S^2)$ -normalized eigenfunctions $\Phi_{l,m}^{\omega}$ can be directly expressed in the spherical harmonics as given in Eq. (3.41):*

$$\Phi_{l,m}^{\omega}(\mathbf{n}^{\omega}(\beta, \gamma)) = \sum_{j=0}^{\infty} \frac{d_j^{l,m}}{\|d^{l,m}\|} Y^{|m|+j,m}(\beta, \gamma), \quad (3.43)$$

with coefficients $d_j^{l,m} = d_j^{l,m}(\rho)$ depend only on $\rho = \sqrt{\frac{D_{33}}{D_{44}} \|\omega\|}$, and are given by Eq. (A.9) in Appendix A.1.

3.2.6 Time-integrated processes and resolvent kernels

Under the assumption of exponentially distributed traveling time, the probability of finding a particle at a certain position and orientation can be expressed in terms of the resolvent $(Q_1 - \alpha I)^{-1}$ of the generator Q_1 , recall Eq. (3.2) and Section 2.2.3. It can be seen that for eigenfunctions of B_{ω}^1 , we have:

$$(B_{\omega}^1 - \alpha I)^{-1} \Phi_{l,m}^{\omega} = \frac{1}{\lambda_r^{l,m} - \alpha} \Phi_{l,m}^{\omega}. \quad (3.44)$$

It follows that the resolvent kernel is given by

$$\hat{R}_\alpha^1(\boldsymbol{\omega}, \mathbf{n}) := -\alpha ((\mathcal{B}_\omega^1 - \alpha I)^{-1} \delta_{\mathbf{e}_z}) = \alpha \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{1}{\alpha - \lambda_r^{l,m}} \overline{\Phi_{l,m}^\omega(\mathbf{e}_z)} \Phi_{l,m}^\omega(\mathbf{n}). \quad (3.45)$$

Thereby the probability density of finding a random walker at a certain position \mathbf{y} with orientation \mathbf{n} , regardless of the traveling time, is given by:

$$\begin{aligned} P_\alpha^1(\mathbf{y}, \mathbf{n}) &= \left(\mathcal{F}_{\mathbb{R}^3}^{-1} (\hat{R}_\alpha^1) *_{\mathbb{R}^3 \times S^2} U \right) (\mathbf{y}, \mathbf{n}) \\ &= \left(\mathcal{F}_{\mathbb{R}^3}^{-1} \left(\boldsymbol{\omega} \mapsto \alpha \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{1}{\alpha - \lambda_r^{l,m}} \overline{\Phi_{l,m}^\omega(\mathbf{e}_z)} \Phi_{l,m}^\omega(\cdot) \right) *_{\mathbb{R}^3 \times S^2} U \right) (\mathbf{y}, \mathbf{n}). \end{aligned} \quad (3.46)$$

3.2.7 From the hypo-elliptic diffusion kernel to the elliptic diffusion kernel

So far in this section we have restricted our diffusion by choosing only D_{33} and $D_{44} = D_{55}$ as nonzero entries in \mathbf{D} , recall (2.13), motivated from the use of this process in applications. However, even for elliptic diffusion it is still possible to obtain exact solutions, with just a simple transformation from the hypo-elliptic case. We are still required to use only legal generators, in the sense discussed in Section 2.2.2. Furthermore, for all differentiable functions $\tilde{U} : SE(3) \rightarrow \mathbb{R}$ induced by $U : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}$ via (2.5) one has $\mathcal{A}_6 \tilde{U} = (\mathcal{A}_6)^2 \tilde{U} = 0$.

Hereby, also the case of elliptic diffusion can be considered:

$$\mathbf{D} = \text{diag}(D_{11}, D_{11}, D_{33}, D_{44}, D_{44}, 0), \quad D_{11}, D_{33}, D_{44} > 0.$$

The generator \tilde{Q}_E of the evolution on $SE(3)$ and the generator Q_E on $\mathbb{R}^3 \times S^2$ then become:

$$\begin{aligned} \tilde{Q}_E &:= D_{11}(\mathcal{A}_1^2 + \mathcal{A}_2^2) + D_{33}\mathcal{A}_3^2 + D_{44}(\mathcal{A}_4^2 + \mathcal{A}_5^2), \\ Q_E &:= D_{11} \|\mathbf{n} \times \nabla\|^2 + D_{33}(\mathbf{n} \cdot \nabla)^2 + D_{44} \Delta_{S^2} \\ &= D_{11} \|\mathbf{n} \times \nabla\|^2 + Q_1. \end{aligned} \quad (3.47)$$

Recall Remark 2.2.2 for the relation between the group and quotient generators. The operator \mathcal{B}_ω on $\mathbb{H}_2(S^2)$, obtained as before from applying a Fourier transform in the spatial variables, changes accordingly:

$$\mathcal{B}_\omega^E = \mathcal{B}_\omega^1 - D_{11}r^2 \sin^2 \beta = D_{44}\Delta_{S^2} - (D_{33} - D_{11})r^2 \cos^2 \beta - r^2 D_{11}. \quad (3.48)$$

It is now fairly straightforward to obtain from Theorem 3.2.3 the following corollary:

Corollary 3.2.5. *Let $D_{33} > D_{11} > 0$ and $t > 0$. Then the elliptic heat kernel $K_t^E = e^{tQ_E} \delta_{(\mathbf{0}, \mathbf{e}_z)}$ is given by*

$$K_t^E(\mathbf{y}, \mathbf{n}) = \mathcal{F}_{\mathbb{R}^3}^{-1} \left(\omega \mapsto \sum_{l=0}^{\infty} \sum_{m=-l}^l \overline{\Phi_{l,m}^{E,\omega}(\mathbf{e}_z)} \Phi_{l,m}^{E,\omega}(\mathbf{n}) e^{(\lambda_r^{l,m} - \|\omega\|^2 D_{11})t} \right) (\mathbf{y}). \quad (3.49)$$

with $\Phi_{l,m}^{E,\omega}(\mathbf{n}^\omega(\beta, \gamma)) = S^{l,m} \frac{1}{\rho \sqrt{\frac{D_{33}-D_{11}}{D_{33}}}} (\cos \beta) C_m(\gamma)$. For $D_{11} \downarrow 0$, one recovers the hypo-elliptic diffusion kernel K_t^1 computed in Theorem 3.2.3.

Proof. Recall that $\rho = \sqrt{\frac{D_{33}}{D_{44}}} \|\omega\| = \sqrt{\frac{D_{33}}{D_{44}}} r$. Then the result follows by (3.48) and the transformation

$$r \mapsto \sqrt{\frac{D_{33} - D_{11}}{D_{33}}} r \text{ and } \lambda \mapsto \lambda - r^2 D_{11}.$$

Finally, the limit $D_{11} \downarrow 0$ can be interchanged with the sum in the series, since by application of the Weierstrass criterion (and the existence of a uniform bound on all eigenfunctions $\Phi_{l,m}^{E,\omega}(\mathbf{n})$) the series is uniformly converging for all $t > 0$. \square

3.3 Derivation of the exact solutions for convection-diffusion kernels on $\mathbb{R}^3 \times S^2$

In this section we consider the second central equation of this chapter, Eq. (3.1) with $i = 2$, for the 3D direction process or convection-diffusion process. We focus here on the time-integrated process, as this has proven to be more useful in applications. We show how exact solutions for the resolvent kernel can be found. Similar to the approach in Section 3.2, we derive eigenfunctions for the corresponding evolution operator in the Fourier domain. However, this operator can no longer be transformed into the standard Sturm-Liouville form. We therefore use the framework of perturbations of self-adjoint operators [Kat76, Mak12] to prove important properties of the eigenvalues and to prove completeness of the eigenfunctions.

The convection-diffusion system that we consider is the following, for $\mathbf{y} \in \mathbb{R}^3, \mathbf{n} \in S^2$:

$$\begin{cases} \partial_t W(\mathbf{y}, \mathbf{n}, t) = Q_2 W(\mathbf{y}, \mathbf{n}, t) = (-\mathbf{n} \cdot \nabla_{\mathbb{R}^3}) + D_{44} \Delta_{S^2} W(\mathbf{y}, \mathbf{n}, t), & t \geq 0, \\ W(\mathbf{y}, \mathbf{n}, 0) = U(\mathbf{y}, \mathbf{n}). \end{cases} \quad (3.50)$$

The equation for the time-integrated, resolvent process is given by:

$$((\mathbf{n} \cdot \nabla_{\mathbb{R}^3}) - D_{44}\Delta_{S^2} - \alpha I) P_\alpha^2(\mathbf{y}, \mathbf{n}) = \alpha U(\mathbf{y}, \mathbf{n}). \quad (3.51)$$

Again we fix $\boldsymbol{\omega} \in \mathbb{R}^3$ and the operator \mathcal{B}_ω^2 (superscript 2) corresponding to Q_2 now becomes:

$$\mathcal{B}_\omega^2 = D_{44}\Delta_{S^2} - (i\boldsymbol{\omega} \cdot \mathbf{n}). \quad (3.52)$$

When we express \mathbf{n} in spherical coordinates β, γ with respect to $\boldsymbol{\omega}$ as was done in (3.25) and Fig. 3.1, the differential operator in β, γ becomes:

$$\mathcal{B}_\omega^2 = \frac{D_{44}}{\sin \beta} \frac{\partial}{\partial \beta} \left(\sin \beta \frac{\partial}{\partial \beta} \right) + \frac{D_{44}}{\sin^2 \beta} \frac{\partial^2}{\partial \gamma^2} - (ir \cos \beta). \quad (3.53)$$

Since the Laplace-Beltrami operator is symmetric and the multiplication operator has a purely imaginary symbol, the operator \mathcal{B}_ω^2 in this case is not symmetric and not self-adjoint, but the adjoint $\mathcal{B}_\omega^{2,*}$ of \mathcal{B}_ω^2 does satisfy:

$$\mathcal{B}_\omega^{2,*} f = \overline{\mathcal{B}_\omega^2 f}, \quad \text{for all } f \in \mathbb{H}_2(S^2). \quad (3.54)$$

Here we note that the domain of the closed unbounded operator \mathcal{B}_ω^2 is the Sobolev space $\mathbb{H}_2(S^2)$, equipped with the $\mathbb{L}_2(S^2)$ -norm. It will turn out to be useful to regard \mathcal{B}_ω^2 as the sum of a self-adjoint operator and a bounded operator $\mathcal{M} : \mathbb{L}_2(S^2) \rightarrow \mathbb{L}_2(S^2)$, that just applies a multiplication, $(\mathcal{M}f)(\mathbf{n}^\omega(\beta, \gamma)) = \cos \beta \cdot f(\mathbf{n}^\omega(\beta, \gamma))$:

$$\mathcal{B}_\omega^2 = D_{44}\Delta_{S^2} - ir \cos \beta = D_{44}\Delta_{S^2} - ir\mathcal{M}, \quad r = \|\boldsymbol{\omega}\|. \quad (3.55)$$

So in particular, as before, the operator $\mathcal{B}_0^2 = D_{44}\Delta_{S^2}$ has the spherical harmonics as eigenfunctions. We denote the eigenfunctions of \mathcal{B}_ω^2 with $\Psi_{l,m}^\omega$ and the corresponding eigenvalues with $\lambda_r^{l,m}$, even though the eigenvalues are not the same as in Section 3.2. Again we assume that the eigenfunctions can be written as the product-form $\Psi_{l,m}^\omega(\mathbf{n}^\omega(\beta, \gamma)) = B(\beta)C(\gamma)$, leading to two ordinary differential equations. The equation with variable γ is the same as before, with solutions $e^{im\gamma}$. The equation for variable β can be written as

$$\frac{1}{\sin \beta} \frac{d}{d\beta} \left(\sin \beta \frac{dB(\beta)}{d\beta} \right) + \left[- \left(\frac{1}{D_{44}} ir \cos \beta + \frac{\lambda_r}{D_{44}} \right) - \frac{m^2}{\sin^2 \beta} \right] B(\beta) = 0, \quad (3.56)$$

i.e., as

$$\frac{1}{\sin \beta} \frac{d}{d\beta} \left(\sin \beta \frac{dB(\beta)}{d\beta} \right) + \left[- \left(i\rho \cos \beta + \tilde{\lambda}_\rho \right) - \frac{m^2}{\sin^2 \beta} \right] B(\beta) = 0, \quad (3.57)$$

now with

$$\rho = \frac{r}{D_{44}}, \quad \tilde{\lambda}_\rho = \frac{\lambda_r}{D_{44}}. \quad (3.58)$$

We define the differential operator \mathcal{B}_ρ^m as

$$\mathcal{B}_\rho^m := \left(\frac{1}{\sin \beta} \frac{d}{d\beta} \left(\frac{1}{\sin \beta} \frac{d}{d\beta} \right) \right) - \frac{m^2}{\sin^2 \beta} - i\rho \cos \beta = \mathcal{B}_0^m - i\rho \mathcal{M}, \quad (3.59)$$

with slight abuse of notation, since now $\mathcal{M} : \mathbb{L}_2([0, \pi]) \rightarrow \mathbb{L}_2([0, \pi])$. Then Eq. (3.57) can be rewritten as:

$$\mathcal{B}_\rho^m B(\beta) = \tilde{\lambda}_\rho B(\beta), \quad \beta \in [0, \pi]. \quad (3.60)$$

We explicitly denote the dependence on ρ and the separation constant m , as we need it later in our spectral analysis of the operator.

3.3.1 The generalized spheroidal wave equation

After applying the transformation $x = \cos \beta$, $y(x) = B(\beta)$ to Eq. (3.57), we get:

$$(1 - x^2) \frac{d^2 y(x)}{dx^2} - 2x \frac{dy(x)}{dx} + \left[-\rho i x - \tilde{\lambda}_\rho - \frac{m^2}{1 - x^2} \right] y(x) = 0, \quad -1 \leq x \leq 1. \quad (3.61)$$

This equation now has the form of a specific case of the generalized spheroidal wave equation (GSWE) [Lea86, OLBC10, Sec. 30.12].

Remark 3.3.1. *In literature [Lea86, Fig07], the generalized spheroidal wave equation also appears in the following form:*

$$x(x - x_0) \frac{d^2 y(x)}{dx^2} + (C_1 + C_2 x) \frac{dy(x)}{dx} + [\omega^2 x(x - x_0) - 2\eta\omega(x - x_0) + C_3] y(x) = 0. \quad (3.62)$$

In this equation C_i , ω , η are constants and the equation has singularities at $x = 0$ and $x = x_0$. With an appropriate choice of constants (taking the limit $\omega \rightarrow 0$, such that $2\eta\omega$ stays bounded and nonzero) the GSWE can be brought to the form of Equation (3.61).

We refer to Appendix A.2 for the derivation of the eigenvalues $\tilde{\lambda}_\rho^{l,m}$ of the GSWE and the corresponding eigenfunctions that we denote with $GS_\rho^{l,m}$. Here we just state that the eigenfunctions of \mathcal{B}_ω^2 are given by

$$\Psi_{l,m}^\omega(\mathbf{n}^\omega(\beta, \gamma)) = GS_\rho^{l,m}(\cos \beta) \frac{e^{im\gamma}}{\sqrt{2\pi}}, \quad l \in \mathbb{N}_0, m \in \mathbb{Z}, |m| \leq l, \rho = \frac{\|\omega\|}{D_{44}}, \quad (3.63)$$

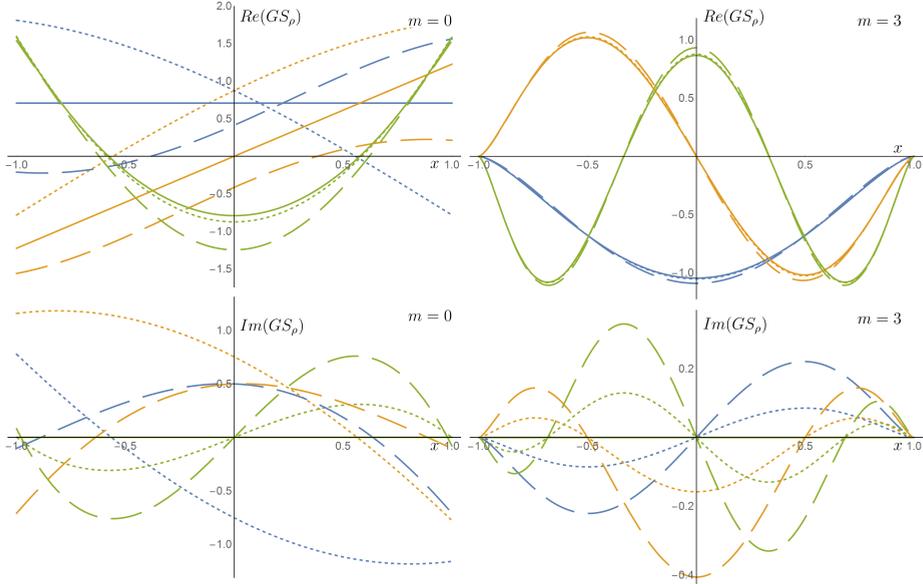


Figure 3.4: Plot of the real (top) and imaginary (bottom) part of the generalized spheroidal wave functions $GS_\rho^{l,m}(x)$ for $m = 0$ (left) and $m = 3$ (right), with $\rho = 0, 2, 5$ (indicated with solid, small dashed and long dashed lines, respectively) and $l = m, m + 1, m + 2$ (indicated with blue, yellow and green, respectively).

in which we used the same substitution $x = \cos \beta$ as before. The functions $GS_\rho^{l,m}$ for certain ρ , l and m are shown in Fig. 3.4. Recall (3.58) for the relation $\lambda_r^{l,m} = -D_{44} \tilde{\lambda}_\rho^{l,m}$ between the eigenvalues corresponding to $\Psi_{l,m}^\omega$ and $GS_\rho^{l,m}$, respectively.

From property (3.54) the following can be derived:

$$\begin{aligned} \lambda_r^{l,m} (\Psi_{l,m}^\omega, \overline{\Psi_{l',m'}^\omega})_{\mathbb{L}_2(S^2)} &= (\mathcal{B}_\omega^2 \Psi_{l,m}^\omega, \overline{\Psi_{l',m'}^\omega})_{\mathbb{L}_2(S^2)} \stackrel{(3.54)}{=} (\Psi_{l,m}^\omega, \overline{\mathcal{B}_\omega^2 \Psi_{l',m'}^\omega})_{\mathbb{L}_2(S^2)} \\ &= (\Psi_{l,m}^\omega, \overline{\lambda_r^{l',m'} \Psi_{l',m'}^\omega})_{\mathbb{L}_2(S^2)} = \lambda_r^{l',m'} (\Psi_{l,m}^\omega, \overline{\Psi_{l',m'}^\omega})_{\mathbb{L}_2(S^2)}. \end{aligned} \quad (3.64)$$

This implies that

$$\lambda_r^{l,m} = \lambda_r^{l',m'} \vee (\Psi_{l,m}^\omega, \overline{\Psi_{l',m'}^\omega})_{\mathbb{L}_2(S^2)} = 0. \quad (3.65)$$

As a result, we see that if $\{\Psi_{l,m}^\omega\}$ is complete and it admits a reciprocal basis $\{\Psi_\omega^{l,m}\}$ such that

$(\Psi_\omega^{l,m}, \Psi_{l',m'}^\omega) = \delta_{l'}^l \delta_{m'}^m$, then for the reciprocal basis functions we have $\Psi_\omega^{l,m} = \overline{\Psi_{l,m}^\omega}$. In the next section, this completeness is further discussed.

3.3.2 Analysis of the time-integrated process

The kernel for the time-integrated process in the spatial Fourier domain corresponds to:

$$\hat{R}_\alpha^2(\boldsymbol{\omega}, \mathbf{n}) := -\alpha ((\mathcal{B}_\omega^2 - \alpha I)^{-1} \delta_{e_z})(\mathbf{n}) = \alpha \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{1}{\alpha - \lambda_r^{l,m}} \frac{\overline{\Psi_{l,m}^\omega(\mathbf{e}_z)} \Psi_{l,m}^\omega(\mathbf{n})}{(\Psi_{l,m}^\omega, \overline{\Psi_{l,m}^\omega})}. \quad (3.66)$$

However, there are conditions on the convergence of this series expression. Since operator \mathcal{B}_ω^2 , in contrast to \mathcal{B}_ω^1 , is no longer self-adjoint, the standard Sturm-Liouville theory, that ensures completeness of the eigenfunctions with negative, real eigenvalues, cannot be applied. In the following we formulate a lemma, on the eigenvalues of \mathcal{B}_ρ^m , and a theorem, on the eigenfunctions, that combined imply that the convergence holds almost everywhere. Only for particular radii $\|\boldsymbol{\omega}\| = D_{44}\rho_n^m$, for some ρ_n^m in the frequency domain there is no convergence, but it can be shown that for any m this happens only on a countable set of $\{\rho_n^m\}_{n \in \mathbb{N}_0}$ that has no accumulation point. As a result, the series in (3.66) converges almost everywhere in the Fourier domain, and thereby the inverse Fourier transform, similar to Eq. (3.40) in Theorem 3.2.3, is still well-defined.

In the next lemma we prove properties of the eigenvalues $\lambda_r^{l,m}$ of the operator \mathcal{B}_ω^2 that are necessary to have convergence of the series in (3.66). For this we also need to consider the operator \mathcal{B}_ρ^m for fixed m , recall the definition in (3.59).

Lemma 3.3.2 (Eigenvalues of the operator \mathcal{B}_ω^2). *We have the following properties for the eigenvalues of \mathcal{B}_ω^2 :*

1. *Let $m \in \mathbb{Z}$, then there exists a $\rho_*^m > 0$ such that \mathcal{B}_ρ^m has real eigenvalues. Moreover, there is at most a countable set $\{\rho_n^m\}_{n=1}^{\infty}$ where two eigenvalues collide and branch into a complex conjugate pair of eigenvalues.*
2. *For all $r \geq 0$ the real part of the eigenvalues of \mathcal{B}_ω^2 is negative.*

Proof. We prove the two points subsequently:

1. For $\rho = 0$, the operator \mathcal{B}_0^m , recall (3.59), is self-adjoint, negative semi-definite and therefore all eigenvalues $\tilde{\lambda}_0^{l,m} = -l(l+1)$, $l \geq |m|$ are real, negative and simple. From the spectral inclusion theorem [Ves07] it follows for the spectrum $\sigma(\mathcal{B}_\rho^m)$ that:

$$\sigma(\mathcal{B}_\rho^m) \subset \{\lambda \in \mathbb{C} \mid \text{dist}(\lambda, \sigma(\mathcal{B}_0^m)) \leq \|i\rho\mathcal{M}\|\}. \quad (3.67)$$

The operator norm is $\|i\rho\mathcal{M}\| = \rho$ and for \mathcal{B}_0^m the minimal distance between two eigenvalues is the distance between the two smallest eigenvalues, when $l = |m|$ and $l = |m| + 1$, resulting in $|\tilde{\lambda}_0^{|m|+1,|m|} - \tilde{\lambda}_0^{|m|,|m|}| = 2(|m| + 1) > 0$. Therefore, we

choose $\rho_*^m = (|m| + 1)$ to guarantee that $\tilde{\lambda}_\rho^{|m|+1,|m|} \neq \tilde{\lambda}_\rho^{|m|,|m|}$ for all $\rho < \rho_*^m$. It can be observed from Eq. (3.61) that when $y(x)$ is an eigenfunction for λ , that $\overline{y(-x)}$ is an eigenfunction for $\bar{\lambda}$. It cannot happen that branching of eigenvalues occurs without two eigenvalues colliding, since the multiplicity of $\lambda(\rho)$ depends continuously on ρ . Since we have shown that no eigenvalues can collide for $\rho < \rho_*^m$, we are guaranteed to have real eigenvalues in this case.

Now according to [MS54], there exists a nonzero analytical function $F(\lambda, \rho)$, such that the equation $F(\lambda, \rho) = 0$ defines the eigenvalues λ_i^m , m fixed, as functions of ρ . We define ρ_n^m to be those values for ρ , in increasing order, for which $\lambda_i^m(\rho) = \lambda_j^m(\rho)$ for some $i \neq j$. Due to the analyticity of F , the set $\{\rho_n^m\}_{n=0}^\infty$ is countable and cannot have an accumulation point [MS54]. We specify the analytic function whose zeros provides for given $m \in \mathbb{Z}$ the values $(\rho_n^m)_{n=0}^\infty$ later (in (Eq. 3.69)).

2. To show that all eigenvalues have a negative real part, it is sufficient to show that the symmetric part of the operator \mathcal{B}_ω^2 is negative definite. Indeed for all $f \in \mathbb{H}_2(S^2)$ (dense in $\mathbb{L}_2(S^2)$), we have:

$$\begin{aligned} \left(\frac{\mathcal{B}_\omega^2 + (\mathcal{B}_\omega^2)^*}{2} f, f \right) &= \frac{1}{2} ((D_{44}\Delta_{S^2} - ri\mathcal{M})f + (D_{44}\Delta_{S^2} + ri\mathcal{M})f, f) \\ &= (D_{44}\Delta_{S^2} f, f) \leq 0. \end{aligned} \quad (3.68)$$

□

The dependency of the eigenvalues on ρ is displayed for two different values of m in Fig. 3.5. In the figure, the points ρ_n^m for which \mathcal{B}_ρ^m has two colliding eigenvalues are indicated with red dots. The points ρ_n^m are in fact zeros of the analytic function

$$\rho \mapsto (\Psi_{l,m}^\omega, \overline{\Psi_{l,m}^\omega}) = \int_{S^2} (\Psi_{l,m}^\omega(\mathbf{n}))^2 d\sigma(\mathbf{n}), \quad (3.69)$$

where the right hand side only depends on $\rho = D_{44}\|\omega\|$. Moreover, for the behavior of the eigenvalues we have

$$\text{Im}(\lambda_{\omega}^{l,m}) = \frac{1}{2} \frac{((i\mathcal{B}_\omega^2 - i(\mathcal{B}_\omega^2)^*)\Psi_{l,m}^\omega, \Psi_{l,m}^\omega)}{(\Psi_{l,m}^\omega, \Psi_{l,m}^\omega)} = r \frac{(\mathcal{M}\Psi_{l,m}^\omega, \Psi_{l,m}^\omega)}{(\Psi_{l,m}^\omega, \Psi_{l,m}^\omega)} = \mathcal{O}(r). \quad (3.70)$$

In particular, for m fixed and $\rho \leq \rho_0^m$, all eigenvalues are real, and hence $(\mathcal{M}\Psi_{l,m}^\omega, \Psi_{l,m}^\omega) = 0$. We use Lemma 3.3.2 in the next theorem, which proves the solution of the time-integrated differential equation is unique.

Theorem 3.3.3. *Let $\omega \in \mathbb{R}^3$ be given. Let $\alpha > 0$. Then*

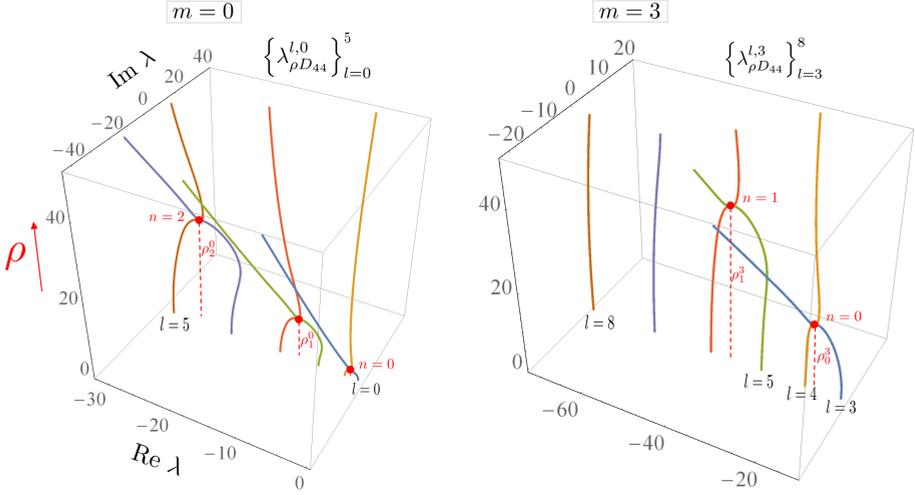


Figure 3.5: Plot of the real and imaginary parts of the first 6 eigenvalues of \mathcal{B}_ρ^m for $m = 0$ (left) and $m = 3$ (right). Note that all eigenvalues are real for sufficiently small ρ . When ρ increases, each time two eigenvalues collide and branch into two complex conjugate eigenvalue pairs. Comparing the left and right figure, we note that the higher m , the higher the values for ρ where this branching occurs. Moreover, we have that $\text{Im}(\lambda_{\rho D_{44}}) \sim \mathcal{O}(\rho D_{44})$.

1. (Existence of the resolvent) the resolvent operator $(\mathcal{B}_\omega^2 - \alpha I)^{-1}$ exists, i.e., the unbounded operator $(\mathcal{B}_\omega^2 - \alpha I) : \mathbb{H}_2(S^2) \rightarrow \mathbb{L}_2(S^2)$ is invertible.
2. (Completeness) there exists a complete basis of generalized eigenfunctions of the operator \mathcal{B}_ω^2 . If $\|\omega\| \neq D_{44}\rho_n^m$, these generalized eigenfunctions are true eigenfunctions and coincide with the eigenfunctions $\Psi_{l,m}^\omega$ derived above. So for $\|\omega\| \neq D_{44}\rho_n^m$ the resolvent operator $(\mathcal{B}_\omega^2 - \alpha I)^{-1}$ is diagonalizable.

Proof. 1. (Existence of the resolvent) Let $\omega \in \mathbb{R}^3$ and $\alpha > 0$ given, $r = \|\omega\|$. Injectivity of $(\mathcal{B}_\omega^2 - \alpha I)$ follows from the fact that the resolvent operator is bounded from below. For $f \in \mathbb{H}_2(S^2)$, we have:

$$\begin{aligned} ((\mathcal{B}_\omega^2 - \alpha I)f, (\mathcal{B}_\omega^2 - \alpha I)f) &= ((D_{44}\Delta_{S^2} + ir\mathcal{M})f, (D_{44}\Delta_{S^2} + ir\mathcal{M})f) \\ &\quad - 2D_{44}\alpha(\Delta_{S^2}f, f) + \alpha^2(f, f) \geq \alpha^2(f, f) \geq 0. \end{aligned} \quad (3.71)$$

Hence $(\mathcal{B}_\omega^2 - \alpha I)f = 0 \implies f = 0$.

To show surjectivity, we start by noting that in general,

$$(\mathcal{R}(\mathcal{B}_\omega^2 - \alpha I))^\perp = \mathcal{N}((\mathcal{B}_\omega^2)^* - \alpha I).$$

Now let $f \in (\mathcal{R}(\mathcal{B}_\omega^2 - \alpha I))^\perp$ and $f \in \mathbb{H}_2(S^2)$, then

$$(\mathcal{B}_\omega^2)^* f - \alpha f = 0 \iff \overline{\mathcal{B}_\omega^2 f} - \alpha f = 0 \iff \mathcal{B}_\omega^2 \bar{f} = \alpha \bar{f} \iff (\mathcal{B}_\omega^2 - \alpha I) \bar{f} = 0,$$

but injectivity of $\mathcal{B}_\omega^2 - \alpha I$ then implies that $\bar{f} = 0 = f$. It follows that $(\mathcal{R}(\mathcal{B}_\omega^2 - \alpha I))^\perp$ equals $\{0\}$ and because of closedness of both \mathcal{B}_ω^2 and I , the surjectivity follows from the closed range theorem [Yos80]. Hence $(\mathcal{B}_\omega^2 - \alpha I)^{-1}$ exists.

2. (Completeness) We first consider the operator \mathcal{B}_ρ^m . By direct computation it can be shown that the multiplication operator $\rho \mathcal{M}$ is bounded, with ρ as a bound for the operator norm. For $\rho = 0$, \mathcal{B}_0^m has simple eigenvalues. Thereby, according to Kato [Kat76, Ch. V, Sect. 5, Th. 4.15a] there exists a complete basis of generalized eigenfunctions. Then \mathcal{B}_ρ^m is closed with compact resolvent $(\mathcal{B}_\rho^m - \alpha I)^{-1}$.

In the case of $\rho \neq \rho_n^m$, we still need to show that the basis of generalized eigenfunctions correspond to the actual eigenfunctions $\Psi_{l,m}^\omega$ as computed above. For $\|\omega\| < D_{44}\rho^*$ this is clear. For $\rho \geq \rho^*$, it follows by analytic extension in ρ , which is exactly what happens when we write down the eigenfunctions as a series of Legendre functions as in (A.16), as this boils down to a Taylor series in ρ .

Furthermore, the functions $GS_\rho^{l,m}$ are uniformly bounded on $[-1, 1]$ and thereby form a Riesz basis [Mak12], which makes the reciprocal basis unique. The reciprocal basis has the property that

$$(\Psi_\omega^{l,m}, \Psi_{l',m'}^\omega) = \delta_{l'}^l \delta_{m'}^m, \quad l, l' \in \mathbb{N}_0, \quad m, m' \in \mathbb{Z}, \quad |m| \leq l, \quad |m'| \leq l'. \quad (3.72)$$

In fact we have $\Psi_\omega^{l,m} = \mathcal{S}^{-1} \Psi_{l,m}^\omega$, where $\mathcal{S} : \mathbb{L}_2(S^2) \rightarrow \mathbb{L}_2(S^2)$ denotes the frame operator, given by

$$\mathcal{S}f = \sum_{l=0}^{\infty} \sum_{m=-l}^l (f, \Psi_{l,m}^\omega) \Psi_{l,m}^\omega. \quad (3.73)$$

From the properties (3.54) and (3.65) it follows that the reciprocal basis $\Psi_\omega^{l,m}$ of $\Psi_{l,m}^\omega$ is linearly proportional to the conjugate basis, which implies that the reciprocal basis $\{\Psi_\omega^{l,m}\}$ is complete. Therefore, for any $f \in \mathbb{L}^2(S^2)$, there is the convergent series representation

$$(\mathcal{B}_\omega^2 - \alpha I)^{-1} f = \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{1}{\lambda_r^{l,m} - \alpha} \frac{(f, \overline{\Psi_{l,m}^\omega}) \Psi_{l,m}^\omega}{(\Psi_{l,m}^\omega, \overline{\Psi_{l,m}^\omega})}, \quad D_{44} \|\omega\| \neq \rho_n^m. \quad (3.74)$$

Hence for $\rho \neq \rho_n^m$, $(\mathcal{B}_\omega^2 - \alpha I)^{-1}$ is diagonalizable with eigenfunctions $\Psi_{l,m}^\omega$ and eigenvalues $1/(\lambda_r^{l,m} - \alpha)$ with strictly negative real part. \square

3.3.2.1 Main theorem

The following theorem summarizes the result regarding eigenfunctions of the operator \mathcal{B}_ω^2 corresponding to the generator Q_2 :

Theorem 3.3.4. *The eigenfunctions $\Psi_{l,m}^\omega$ of the operator \mathcal{B}_ω^2 in (3.52) are given by:*

$$\Psi_{l,m}^\omega(\beta, \gamma) = GS_\rho^{l,m}(\cos \beta)C_m(\gamma). \quad (3.75)$$

Here $GS_\rho^{l,m}$ is the eigenfunction, given by (A.16), of the generalized spheroidal wave equation. For almost every $\omega \in \mathbb{R}^3$, these eigenfunctions form a complete bi-orthogonal system. Therefore the solution of the convection-diffusion equation on $\mathbb{R}^3 \times S^2$ is given by:

$$P_\alpha^2(\mathbf{y}, \mathbf{n}) = (R_\alpha^2 *_{\mathbb{R}^3 \times S^2} U)(\mathbf{y}, \mathbf{n}), \quad (3.76)$$

with

$$R_\alpha^2(\mathbf{y}, \mathbf{n}) = \mathcal{F}_{\mathbb{R}^3}^{-1} \left(\omega \mapsto \sum_{l=0}^{\infty} \sum_{m=-l}^l \frac{\alpha}{\alpha - \lambda_r^{l,m}} \frac{\overline{\Psi_{l,m}^\omega(\mathbf{e}_z)} \Psi_{l,m}^\omega(\mathbf{n})}{(\Psi_{l,m}^\omega, \overline{\Psi_{l,m}^\omega})} \right) (\mathbf{y}), \quad (3.77)$$

where the series converges in $\mathbb{L}_2(\mathbb{R}^3 \times S^2)$ -sense. With $\lambda_r^{l,m} = -D_{44}l(l+1) + \mathcal{O}(r)$ we denote the countable eigenvalues of \mathcal{B}_ω^2 , with $\|\omega\| = r = \rho D_{44}$. The eigenvalues are disjoint for $\rho \neq 0$, and $\rho \neq \rho_n^m$.

Remark 3.3.5. • For all $\omega \in \mathbb{R}^3$, including the cases $\|\omega\| = D_{44}\rho_n^m$, we can decompose the resolvent into a complete basis of generalized eigenfunctions. In fact the projection onto the generalized eigenspace $E_{\tilde{\lambda}_\rho^{l,m}}$ is given by

$$P_{E_{\tilde{\lambda}_\rho^{l,m}}} = \frac{1}{2\pi i} \oint_J (zI - \mathcal{B}_\rho^m)^{-1} dz, \quad (3.78)$$

with J a Jordan curve enclosing only $\tilde{\lambda}_\rho^{l,m}$ in positive direction, where we recall definition (3.59).

- Now if $\rho \neq \rho_n^m$, we have one-dimensional eigenspaces, spanned by $GS_\rho^{l,m}$:

$$E_{\tilde{\lambda}_\rho^{l,m}} = \text{span} \{GS_\rho^{l,m}\} = \mathcal{N}(\mathcal{B}_\rho^m - \tilde{\lambda}_\rho^{l,m}I). \quad (3.79)$$

In the case that $\rho = \rho_n^m$, we have instead:

$$E_{\tilde{\lambda}_\rho^{l,m}} = \text{span} \left\{ \mathcal{N}(\mathcal{B}_\rho^m - \tilde{\lambda}_\rho^{l,m} I), \mathcal{N}(\mathcal{B}_\rho^m - \tilde{\lambda}_\rho^{l,m} I)^2 \right\}. \quad (3.80)$$

At $\rho = \rho_n^m$, the algebraic multiplicity of $\tilde{\lambda}_\rho^{l,m}$ in $\mathcal{B}_\rho^m|_{E_{\tilde{\lambda}_\rho^{l,m}}}$ is 2, whereas the geometric multiplicity is 1.

3.3.2.2 Time integration with Gamma-distributed traveling times

The kernel for the time-integrated process with exponentially distributed traveling time has a singularity at the origin $(\mathbf{y}, \mathbf{n}) = (\mathbf{0}, \mathbf{e}_z)$. It is possible to derive a relation between the resolvent kernel and the process with Gamma-distributed traveling times, that does not suffer from this singularity. This is beneficial for practical applications, such as completion fields [DDHCG12, Fig. 20]. Thanks to the exact solution representation of the kernels for both processes, we obtain the following refinement and generalization of [DDHCG12, Thm. 12].

Theorem 3.3.6. *The kernel of the time-integrated diffusion ($i = 1$) and convection-diffusion ($i = 2$) process with the assumption of Γ -distributed evolution time T , i.e., $T \sim \Gamma(k, \alpha)$, $k \in \mathbb{N}$, with $E(T) = \frac{k}{\alpha}$, is related to repeated convolutions of the resolvent kernel as follows:*

$$\int_0^\infty K_t^i \Gamma(t; k, \alpha) dt = R_\alpha^i *_{\mathbb{R}^3 \times S^2}^{(k-1)} R_\alpha^i, \quad (3.81)$$

where $\Gamma(t; k, \alpha)$ is the pdf of T . For the case $i = 1$, this kernel does not have a singularity in the origin when $k \geq 2$. For the case $i = 2$, this holds when $k \geq 4$.

Proof. We refer to Appendix A.3 for the proof. □

When comparing kernels for varying k , while keeping the expected value of the traveling time T fixed, using a k higher than the bounds given in the theorem above gives better shaped kernels in practice, with more outward mass and dampened singularities. We use this idea in the visualization of a time-integrated contour completion kernel in Section 3.4, Fig. 3.7.

3.4 Matrix representation of the evolution and resolvent in a Fourier basis

In [ZDStHR16, DvA08] a connection between the exact solutions and a numerical algorithm proposed by August and Zucker in [Aug01, AZ03] was established for the $SE(2) =$

$\mathbb{R}^2 \times S^1$ case. It was mentioned in [DvA08] that the algorithm applies to both contour completion and contour enhancement, and directly relates to their exact solutions and the ones presented in [DF10b]. In this algorithm the Fourier transform on $\mathbb{L}_2(\mathbb{R}^2)$ and $\mathbb{L}_2(S^1)$ are applied subsequently. Here we again establish such a connection between the exact solutions and such a numerical algorithm in the 3D case.

We have seen in the previous sections that the eigenfunctions $\Phi_{l,m}^\omega$ and $\Psi_{l,m}^\omega$ are closely related to the spherical harmonic functions. Using the Fourier transform on $\mathbb{L}_2(S^2)$, we naturally obtain an ordinary differential equation in terms of spherical harmonic coefficients. In Section 3.4.1 we give a derivation of this ODE for the diffusion case, and state the result for the convection-diffusion case.

At the end of Section 3.4.1 we include two remarks: one that shows that in deriving the ODEs we encounter a matrix representation of the evolution operator \mathcal{B}_ω^i and its resolvent, and one that makes the connection with the Fourier transform on $SE(3)$ as presented in Appendix A.4. Finally, we show in Section 3.4.2 that with the procedure presented in this section it is straightforward to compute a numerical solution to the PDEs, since it only requires truncation of the order of the spherical harmonics.

3.4.1 Spherical harmonic expansions of the solutions to (convection-)diffusion equations

Recall that after a Fourier transform in the spatial coordinates, we have the following system:

$$\begin{cases} \partial_t \hat{W}(\omega, \mathbf{n}, t) = (D_{44} \Delta_{S^2} - D_{33} (\omega \cdot \mathbf{n})^2) \hat{W}(\omega, \mathbf{n}, t), & t \geq 0, \\ \hat{W}(\omega, \mathbf{n}, 0) = \hat{U}(\omega, \mathbf{n}). \end{cases} \quad (3.82)$$

Since the spherical harmonics form a basis for $\mathbb{L}_2(S^2)$, we can expand $\hat{W}(\omega, \mathbf{n}, t)$ for fixed ω and t in the spherical harmonic basis. Instead of doing this with standard spherical harmonics, we use a specific type of reoriented spherical harmonics, that we define as follows:

$$Y_\omega^{l,m}(\mathbf{n}) := Y_0^{l,m}(\mathbf{R}_{\omega r^{-1}}^T \mathbf{n}) := Y^{l,m}(\beta, \gamma) = \frac{\varepsilon^m}{\sqrt{2\pi}} P_l^m(\cos \beta) e^{im\gamma}, \quad \text{with } \mathbf{n} = \mathbf{n}^\omega(\beta, \gamma), \quad (3.83)$$

and $m \in \mathbb{Z}$, $l \in \mathbb{N}_0$, $|m| \leq l$. Recall Fig. 3.1 and Eq. (3.25). Rotation $\mathbf{R}_{r^{-1}\omega}$ in (3.83) is defined through its matrix

$$\mathbf{R}_{r^{-1}\omega} = \left(\begin{array}{c|c|c} \frac{(\omega \times \mathbf{e}_z) \times \omega}{\|(\omega \times \mathbf{e}_z) \times \omega\|} & \frac{\omega \times \mathbf{e}_z}{\|\omega \times \mathbf{e}_z\|} & r^{-1}\omega \end{array} \right). \quad (3.84)$$

This rotation maps \mathbf{e}_z onto $r^{-1}\boldsymbol{\omega}$ and \mathbf{e}_y onto $\boldsymbol{\omega} \times \mathbf{e}_z / \|\boldsymbol{\omega} \times \mathbf{e}_z\|$, such that for every β, γ we have that

$$\begin{aligned} \mathbf{n}^\omega(\beta, \gamma) &= \mathbf{R}_{r^{-1}\boldsymbol{\omega}} \mathbf{n}^{\mathbf{e}_z}(\beta, \gamma) \\ \mathbf{R}_{r^{-1}\boldsymbol{\omega}}^T \mathbf{R}_{\boldsymbol{\omega} r^{-1}, \gamma} \mathbf{R}_{\frac{\boldsymbol{\omega} \times \mathbf{e}_z}{\|\boldsymbol{\omega} \times \mathbf{e}_z\|}, \beta} \mathbf{R}_{r^{-1}\boldsymbol{\omega}} \mathbf{e}_z &= \mathbf{R}_{\mathbf{e}_z, \gamma} \mathbf{R}_{\mathbf{e}_y, \beta} \mathbf{e}_z. \end{aligned} \quad (3.85)$$

Now for fixed $\boldsymbol{\omega}$ and t we develop $\hat{W}(\boldsymbol{\omega}, \cdot, t)$ in the basis $\{Y_\omega^{l,m} \mid l \in \mathbb{N}_0, m \in \mathbb{Z}, |m| \leq l\}$ as follows:

$$\hat{W}(\boldsymbol{\omega}, \mathbf{n}, t) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \hat{W}^{l,m}(\boldsymbol{\omega}, t) Y_\omega^{l,m}(\mathbf{n}), \quad (3.86)$$

where we define $\hat{W}^{l,m}(\boldsymbol{\omega}, t)$ (and similarly $\hat{U}^{l,m}(\boldsymbol{\omega})$):

$$\hat{W}^{l,m}(\boldsymbol{\omega}, t) := \int_{S^2} \hat{W}(\boldsymbol{\omega}, \mathbf{n}, t) \overline{Y_\omega^{l,m}(\mathbf{n})} d\sigma(\mathbf{n}). \quad (3.87)$$

Our goal is the recursion in Eq. (3.96) for $\hat{W}^{l,m}(\boldsymbol{\omega}, t)$, that we can use to obtain a solution for these coefficients. To this end we start by substituting (3.87) into our differential equation (3.82):

$$\begin{aligned} \sum_{l=0}^{\infty} \sum_{m=-l}^l Y_\omega^{l,m}(\mathbf{n}) \partial_t \hat{W}^{l,m}(\boldsymbol{\omega}, t) &= \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^l (D_{44} \Delta_{S^2} - D_{33} r^2 \cos^2 \beta) Y_\omega^{l,m}(\mathbf{n}) \hat{W}^{l,m}(\boldsymbol{\omega}, t) \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^l (-D_{44} l(l+1) - D_{33} r^2 \cos^2 \beta) Y_\omega^{l,m}(\mathbf{n}) \hat{W}^{l,m}(\boldsymbol{\omega}, t). \end{aligned} \quad (3.88)$$

We aim at rewriting the term $\cos^2 \beta Y^{l,m}(\beta, \gamma)$, see (3.94) below. In [OLBC10] the following identity for Legendre functions is given:

$$x P_l^m(x) = \frac{l-m+1}{2l+1} P_{l+1}^m(x) + \frac{l+m}{2l+1} P_{l-1}^m(x) =: \xi^{l,m} P_{l+1}^m(x) + \nu^{l,m} P_{l-1}^m(x), \quad m \geq 0. \quad (3.89)$$

Using this identity twice, we get:

$$\begin{aligned} x^2 P_l^m(x) &= \frac{(l-m+1)(l-m+2)}{(2l+3)(2l+1)} P_{l+2}^m(x) \\ &\quad + \frac{(2l(l+1) - 2m^2 - 1)}{4l(l+1) - 3} P_l^m(x) + \frac{(l+m-1)(l+m)}{(2l-1)(2l+1)} P_{l-2}^m(x) \\ &=: \zeta^{l,m} P_{l+2}^m(x) + \eta^{l,m} P_l^m(x) + \alpha^{l,m} P_{l-2}^m(x), \quad m \geq 0. \end{aligned} \quad (3.90)$$

$$\begin{aligned}
\partial_t \hat{W}^{l,m}(\boldsymbol{\omega}, t) &= -D_{44}l(l+1)\hat{W}^{l,m}(\boldsymbol{\omega}, t) + \\
&-D_{33}r^2 \left((\mathbf{M}_1^m)^T_{l-2,l} \hat{W}^{l-2,m}(\boldsymbol{\omega}, t) + (\mathbf{M}_1^m)^T_{l,l} \hat{W}^{l,m}(\boldsymbol{\omega}, t) + (\mathbf{M}_1^m)^T_{l+2,l} \hat{W}^{l+2,m}(\boldsymbol{\omega}, t) \right) \\
&= -D_{44}l(l+1)\hat{W}^{l,m}(\boldsymbol{\omega}, t) + \\
&-D_{33}r^2 \left(\zeta^{l-2,m} \hat{W}^{l-2,m}(\boldsymbol{\omega}, t) + \eta^{l,m} \hat{W}^{l,m}(\boldsymbol{\omega}, t) + \alpha^{l+2,m} \hat{W}^{l+2,m}(\boldsymbol{\omega}, t) \right)
\end{aligned} \tag{3.96}$$

In other words, for m fixed, we need to solve the following system:

$$\begin{cases} \partial_t \hat{\mathbf{w}}^m(\boldsymbol{\omega}, t) = -D_{33}r^2 \mathbf{M}_1^m \hat{\mathbf{w}}^m(\boldsymbol{\omega}, t) - D_{44} \boldsymbol{\Lambda}^m \hat{\mathbf{w}}^m(\boldsymbol{\omega}, t), \\ \hat{\mathbf{w}}^m(\boldsymbol{\omega}, 0) = \hat{\mathbf{u}}^m(\boldsymbol{\omega}), \end{cases} \tag{3.97}$$

with $\hat{\mathbf{w}}^m(\boldsymbol{\omega}, t) = (\hat{W}^{l,m}(\boldsymbol{\omega}, t))_{l=|m|}^\infty$, $\hat{\mathbf{u}}^m(\boldsymbol{\omega}, t) = (\hat{U}^{l,m}(\boldsymbol{\omega}, t))_{l=|m|}^\infty$ and $\boldsymbol{\Lambda}^m = \text{diag}(|m|(|m|+1), (|m|+1)(|m|+2), \dots)$.

Conclusion: the result of all prior computations of this subsection is that for the diffusion equation (3.82), the solution

$$W(\mathbf{y}, \mathbf{n}, t) = [\mathcal{F}_{\mathbb{R}^3}^{-1} \hat{W}(\cdot, \mathbf{n}, t)](\mathbf{y}) = (e^{tQ_1} U)(\mathbf{y}, \mathbf{n}) \tag{3.98}$$

can be expanded as in (3.86), resulting in the ODE (3.96) for each m . The solution of this system in matrix-vector form is given by

$$\hat{\mathbf{w}}^m(\boldsymbol{\omega}, t) = \exp\{(-D_{33}r^2 \mathbf{M}_1^m - D_{44} \boldsymbol{\Lambda}^m)t\} \hat{\mathbf{u}}^m(\boldsymbol{\omega}). \tag{3.99}$$

The same idea of substituting directly a series of spherical harmonics into the equation can be used for the resolvent case of the convection-diffusion equation. After the Fourier transform in the spatial coordinates, this equation reads

$$(\alpha I - \mathcal{B}_\omega^2) \hat{W}(\boldsymbol{\omega}, \mathbf{n}) = (\alpha I - (D_{44} \Delta_{S^2} - i(\boldsymbol{\omega} \cdot \mathbf{n}))) \hat{W}(\boldsymbol{\omega}, \mathbf{n}) = \alpha \hat{U}(\boldsymbol{\omega}, \mathbf{n}). \tag{3.100}$$

The solution

$$W(\mathbf{y}, \mathbf{n}) = [\mathcal{F}_{\mathbb{R}^3}^{-1}](\mathbf{y}) = \alpha(\alpha I - Q_2)^{-1} U(\mathbf{y}, \mathbf{n}) \tag{3.101}$$

can be found by using (as before) the expansion $\hat{W}(\boldsymbol{\omega}, \mathbf{n}) = \sum_{l=0}^\infty \sum_{m=-l}^l \hat{W}^{l,m}(\boldsymbol{\omega}) Y_\omega^{l,m}(\mathbf{n})$. Similar computations yield the following solution in matrix-vector form:

$$\hat{\mathbf{w}}^m(\boldsymbol{\omega}) = \alpha(\alpha I + D_{44} \boldsymbol{\Lambda}^m + ir \mathbf{M}_2^m)^{-1} \hat{\mathbf{u}}^m(\boldsymbol{\omega}). \tag{3.102}$$

Here $\hat{\mathbf{w}}^m(\boldsymbol{\omega}) = (\hat{W}^{l,m}(\boldsymbol{\omega}))_{l=|m|}^{\infty}$ and matrix \mathbf{M}_2^m only has non-zero elements on the upper and lower diagonal:

$$(\mathbf{M}_2^m)^T := \mathbf{N}^m \begin{pmatrix} 0 & \xi^{|m|,|m|} & & \mathbf{O} \\ \nu^{|m|+1,|m|} & 0 & \xi^{|m|+1,|m|} & \\ & \nu^{|m|+2,|m|} & 0 & \xi^{|m|+2,|m|} \\ \mathbf{O} & & \ddots & \ddots \end{pmatrix} (\mathbf{N}^m)^{-1}, \quad (3.103)$$

with $\xi^{l,m}$ and $\nu^{l,m}$ as in (3.89) and \mathbf{N}^m as in (3.93).

Remark 3.4.1. *There is a direct connection between the exact solution presented in Theorem 3.2.3/Corollary 3.2.4 and the solution found via Eq. (3.99). In fact, in case of the exact solutions, the generator and thereby also the evolution operator in (3.99), are diagonalized by the solutions of Eq. (A.10) in Appendix A.1. To clarify this observation, We note that Eq. (A.10) is the eigenvector problem corresponding to the eigenfunction problem in Eq. (3.16), while restricting operator \mathcal{B}_1^ω to the span of spherical harmonics $Y^{l,m}$, $l \geq |m|$, for $m \in \mathbb{Z}$ fixed.*

Remark 3.4.2. *Eq. (3.96) follows from Eq. (3.7) by application of the operator:*

$$(\mathcal{F}_{S^2} \otimes \mathbb{1}_{L_2(\mathbb{R}^3)}) \circ (\mathbb{1}_{L_2(\mathbb{R}^3)} \otimes \mathcal{U}_{\mathbf{R}_{\boldsymbol{\omega}_{r-1}}}) \circ (\mathcal{F}_{\mathbb{R}^3} \otimes \mathbb{1}_{L_2(S^2)}), \quad (3.104)$$

that can be roughly formulated as $\mathcal{F}_{S^2} \circ \mathcal{U}_{\mathbf{R}_{\boldsymbol{\omega}_{r-1}}} \circ \mathcal{F}_{\mathbb{R}^3}$, with $\mathcal{U}_{\mathbf{R}_{\boldsymbol{\omega}_{r-1}}}$ the left-regular representation $\mathcal{U}_{\mathbf{R}}\phi(\mathbf{n}) = \phi(\mathbf{R}^T \mathbf{n})$. There is a close connection between this operator and the Fourier transform and irreducible representations on $SE(3)$, see Appendix A.4.

3.4.2 Numerical implementation of the discrete spherical transform

When using the above procedure to compute the kernel, there are two places where the numerics differ from the exact solution: the series of spherical harmonics is truncated and the Fourier transform is carried out discretely. We introduce a parameter l_{max} to indicate the maximal order of spherical harmonics that is taken into account. Furthermore, we take discrete values for $\boldsymbol{\omega}$ on an equidistant cubic grid, say $\boldsymbol{\omega}_{ijk}$, such that for each $\boldsymbol{\omega}_{ijk} \in \mathbb{R}^3$ the component $\hat{\mathbf{w}}^m$ of the solution requires, for the pure-diffusion case, solving the ODE:

$$\begin{cases} \partial_t \hat{\mathbf{w}}^m(\boldsymbol{\omega}_{ijk}, t) = -D_{33} r^2 \mathbf{M}_{1,l_{max}}^m \hat{\mathbf{w}}^m(\boldsymbol{\omega}_{ijk}, t) - D_{44} \boldsymbol{\Lambda}_{l_{max}}^m \hat{\mathbf{w}}^m(\boldsymbol{\omega}_{ijk}, t), \\ \hat{\mathbf{w}}^m(\boldsymbol{\omega}_{ijk}, 0) = \hat{\mathbf{u}}^m(\boldsymbol{\omega}_{ijk}) := \sum_{l=0}^{l_{max}} Y_{\boldsymbol{\omega}_{ijk}}^{l,m}(\mathbf{e}_z) Y_{\boldsymbol{\omega}_{ijk}}^{l,m}(\mathbf{n}), \end{cases} \quad (3.105)$$

with $\hat{\mathbf{w}}^m = (\hat{W}^{l,m})_{l=|m|}^{l_{max}}$ and $\mathbf{M}_{1,l_{max}}^m, \mathbf{\Lambda}_{l_{max}}^m \in \mathbb{R}^{(l_{max}-|m|+1) \times (l_{max}-|m|+1)}$.

In the convection-diffusion, resolvent case, it comes down to solving:

$$(\alpha I + D_{44} \mathbf{\Lambda}_{l_{max}}^m + ir \mathbf{M}_{2,l_{max}}^m) \hat{\mathbf{w}}^m(\boldsymbol{\omega}) = \alpha \hat{\mathbf{u}}^m(\boldsymbol{\omega}), \quad \hat{\mathbf{w}}^m(\boldsymbol{\omega}, t) = (\hat{W}^{l,m}(\boldsymbol{\omega}, t))_{l=|m|}^{l_{max}}, \quad (3.106)$$

for all $\boldsymbol{\omega} = \boldsymbol{\omega}_{ijk}$ on an equidistant grid:

$$\boldsymbol{\omega}_{ijk} = \left(\frac{i\eta\pi}{N}, \frac{j\eta\pi}{N}, \frac{k\eta\pi}{N} \right), \quad i, j, k \in \{-N, \dots, N\}, \eta \in \mathbb{N}, \quad (3.107)$$

where N denotes the number of samples. We use a discrete centered inverse Fourier transform to go back to the $\mathbb{R}^3 \times S^2$ domain.

The result for the diffusion kernel $K_t^1(\boldsymbol{\omega}, \mathbf{n})$ with $t = 2$, $D_{33} = 1$, $D_{44} = 0.1$, $\eta = 8$, $N = 65$ and $l_{max} = 12$ (resulting in 169 spherical harmonic coefficients) is shown in Fig. 3.6. The convection-diffusion kernel $K_t^2(\boldsymbol{\omega}, \mathbf{n})$ for $k = 1$, $D_{44} = 0.5$, $\eta = 4$, $N = 65$, $l_{max} = 12$ is given on the left in Fig. 3.7. Numerically integrating these kernels for different t , using a Γ -distribution with $k = 4$ and $\alpha = 0.25$ gives the result as displayed on the right in Fig. 3.7.

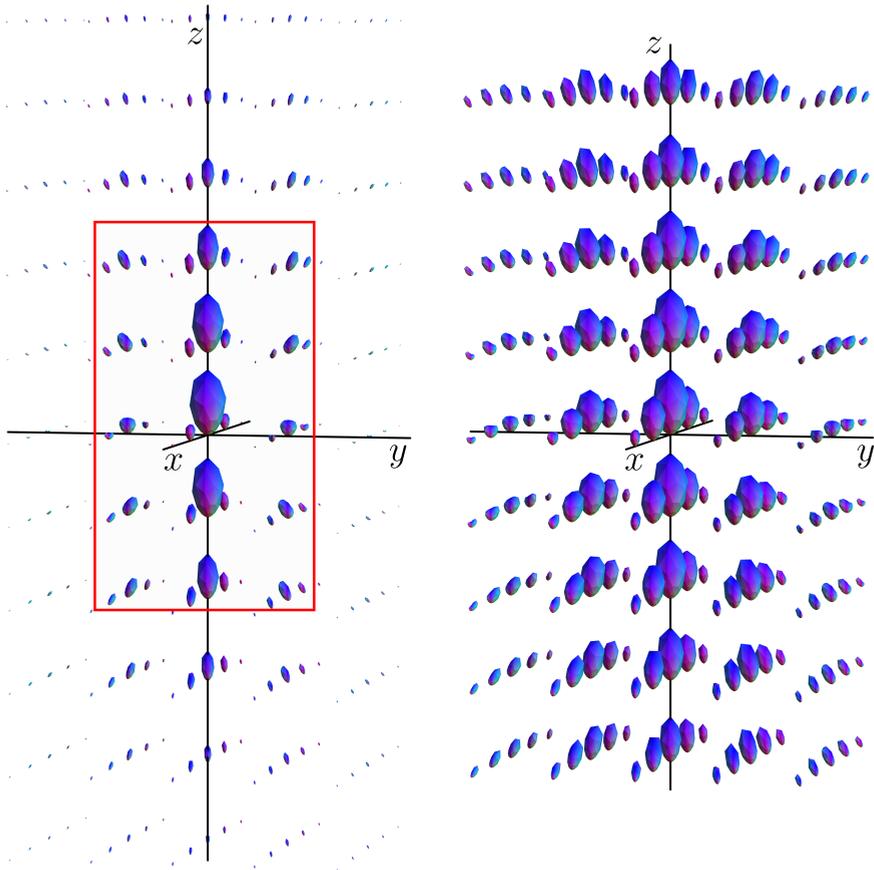


Figure 3.6: Glyph field visualization (as explained in Section 1.1.1) of the kernel $K_{t=2}^1(\boldsymbol{\omega}, \mathbf{n})$, with a higher resolution on the right. For this kernel, $D_{33} = 1$, $D_{44} = 0.1$.

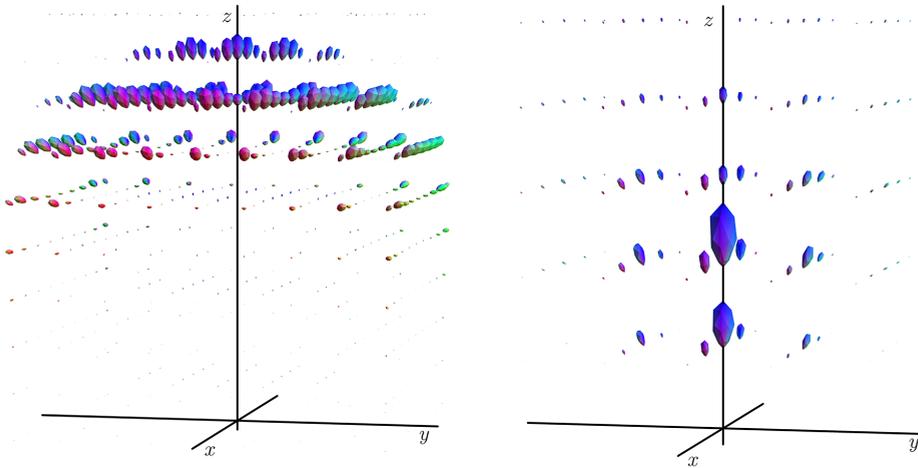


Figure 3.7: Glyph field visualization of the time dependent kernel $K_{t=1}^2$ for the convection-diffusion case, with $D_{44} = 0.5$ (left) and the time-integrated kernel (right), where we used a Gamma-distribution with $k = 4$ and $\alpha = .25$.

3.5 Conclusion

We have provided new explicit solutions for the hypo-elliptic diffusion process for (restricted) Brownian motion on $SE(3)$ and for the convection-diffusion process on $SE(3)$, that is an extension of Mumford's 3D-direction process [Mum94]. Both processes are of interest for dMRI applications as shown in e.g. [MSS13,DDHCG12] (for convection-diffusion) and Chapter 6 (for diffusion).

The solutions were derived by applying a Fourier transform in the spatial variables and a particular (frequency dependent) choice of coordinates, yielding for both processes a separable second order differential operator. Using the spectral decomposition of this operator, we have obtained a series expression for the solution kernel of the evolution equation and the kernel for the time-integrated process, related to the resolvent of this operator.

In the diffusion case, the eigenfunctions encountered in the spectral decomposition are similar to spherical harmonics, but require spheroidal wave functions instead of Legendre functions. Convergence of the series expressions in terms of the eigenfunctions was shown using Sturm-Liouville theory. The final expression for the exact solution is given in Theorem 3.2.3.

For the convection-diffusion case, generalized spheroidal wave functions are needed, that we derive applying a non-standard expansion in Legendre functions. In this case, in contrast to the diffusion case, the considered operator is no longer self-adjoint and as a result, standard Sturm-Liouville cannot be used for proving convergence of the series. Instead we prove completeness of the eigenfunctions and specific properties of the spectrum using perturbation theory of linear (self-adjoint) operators. The exact solution for the resolvent case is given in Theorem 3.3.4.

We have also established a numerical algorithm on $SE(3)$ that generalizes the algorithm in [Aug01,AZ03] for Mumford's 2D direction process to the 3D case. Using this method, approximate solutions can be found by rewriting the equation into an ODE in terms of expansion coefficients w.r.t. a rotated spherical harmonic basis. We connect this numerical algorithm to the exact solutions, showing that it in essence just diagonalizes (the matrix of) the resolvent operator. Truncation of the exact series representation of the kernels yield the same result. Results of the truncated kernels are shown in Figs. 3.6 and 3.7.

Finally, we have algebraically derived the same exact solutions via harmonic analysis on Lie groups, using the Fourier transform on $SE(3)$, see Theorem A.4.3.

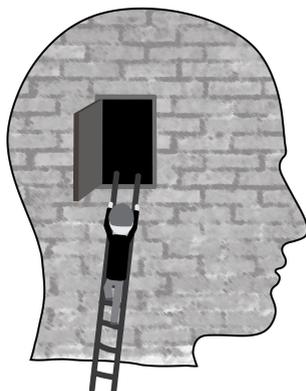
Approximations to Diffusion Kernels on $SE(3)$

Based on:

[PSMD15] J. M. Portegies, G. R. Sanguinetti, S. P. L. Meesters, and R. Duits. New Approximation of a Scale Space Kernel on $SE(3)$ and Applications in Neuroimaging. *SSVM*, number 9087 in LNCS, pages 40–52. Springer International Publishing, May 2015.

[PD17] J. M. Portegies and R. Duits, “New Exact and Numerical Solutions of the (Convection–)Diffusion Kernels on $SE(3)$,” *Differential Geometry and its Applications*, vol. 53, pp. 182–219, Aug. 2017.

[BCP17] E. J. Bekkers, D. Chen, and J. M. Portegies. Nilpotent Approximations of Sub-Riemannian Distances for Fast Perceptual Grouping of Blood Vessels in 2D and 3D. *arXiv:1707.02811 [math]*, July 2017. *Accepted for publication in JMIV, Special Issue ‘Differential Geometry and Orientation Analysis’.*



4.1 Introduction

Just as Chapter 3, this chapter treats the hypo-elliptic diffusion equations on $\mathbb{R}^3 \times S^2$ and $SE(3)$, only this time we look for numerical approximations of the heat kernels. This chapter is merging the papers [BCP17] and [PSMD15]. The paper [PSMD15] focused on a new approximation for the diffusion kernel, and had its origin in the fact that we wanted fast and simple implementations of such PDEs applied to dMRI data. Previous approximations were not satisfactory, for reasons explained later. However, a complete overview of how to derive the kernel was (due to page restrictions) beyond the scope of that paper.

Very recently in [BCP17], for the purpose of perceptual grouping of local orientations for reconstruction of vessel trees, the connection between sub-Riemannian distances and norms on vectors (of so-called coordinates of the first kind) obtained from the logarithmic map on $SE(3)$ was worked out in detail, based on the generic derivations of [RS76, NSW85]. The application of vessel tree reconstruction has solely been studied and described by the first author of [BCP17], and is beyond the scope of this thesis. The theory however complements the approach described in [PSMD15] and therefore the two papers are united here.

The group theoretical approach is roughly as follows. By truncating the Baker-Campbell-Hausdorff (BCH) formula at a certain order (ignoring nested commutators as if the group were nilpotent), we obtain a group product on the coordinates of the first kind. The obtained Heisenberg-type group, which we denote with $(SE(3))_0$, locally approximates $SE(3)$. We then define a norm on $(SE(3))_0$ based on the Folland-Kaplan-Korányi gauge, which is known for its relation to the fundamental solution of the sub-Laplacian on the Heisenberg group [Fol73, Kap80, Kor82]. Hence this approach can be used to approximate the kernel for hypo-elliptic diffusion on $SE(3)$. The distance associated with the Folland-Kaplan-Korányi type norm on $(SE(3))_0$ is locally equivalent to the sub-Riemannian distance on $SE(3)$, as was proven (in a more general setting) in the seminal work by Nagel, Stein and Wainger [NSW85]. Finally, we obtain a new tangible approximation on the group quotient $\mathbb{R}^3 \times S^2$, that has the right symmetries.

Alternatively, because of the elegant (well-known) relation of these diffusion processes to stochastic processes, we also provide a way to approximate the same kernels with a Monte Carlo simulation. The diffusion equation is a Fokker-Planck equation describing the evolution of a probability density function in time. The working principles of such a simulation are well-known: by simulating a large number of random walks, and by dividing the position-orientation space in bins of finite size, the probability density of the end-position of the random walker can easily be obtained, see Fig. 4.1, which approximates the heat kernel. Note that only spatial projections of the paths are given. On

the left, for diffusion/contour enhancement, the random walker can only take an arbitrary step forward or backwards in the direction of its current orientation, and change its orientation with an arbitrary small step. On the right, for convection-diffusion/contour completion, the random walker can change its orientation in the same way with small steps, but always takes a step forward in its current direction. The same approach was part of the comparison of several numerical methods and exact solutions of essentially the same diffusions on $SE(2)$ in [ZDStHR16], and turned out accurate for a sufficiently large number of simulated paths. Although this approach can easily be implemented for both processes, only the diffusion process on $SE(3)$ is further discussed in this chapter.

Sections 2.1 and 2.2 on Lie groups are preliminary to this Chapter, and quite some notation is inherited from those sections. To avoid having to flip to other pages in other sections too often, we repeat the equation of interest for restricted diffusion on position-orientation space:

$$\begin{cases} \partial_t W(\mathbf{y}, \mathbf{n}, t) = D_{33}(\mathbf{n} \cdot \nabla)^2 W(\mathbf{y}, \mathbf{n}, t) + D_{44} \Delta_{LB} W(\mathbf{y}, \mathbf{n}, t), \\ W(\mathbf{y}, \mathbf{n}, 0) = U(\mathbf{y}, \mathbf{n}), \end{cases} \quad (4.1)$$

with $\mathbf{y} \in \mathbb{R}^3$, $\mathbf{n} \in S^2$, $t, D_{33}, D_{44} > 0$, $W : \mathbb{R}^3 \times S^2 \times \mathbb{R} \rightarrow \mathbb{R}$, $U : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}$ and Δ_{LB} the Laplace-Beltrami operator on the sphere. We want to find an approximation for the impulse response or convolution kernel of these PDEs, i.e., the solution with initial condition $U(\mathbf{y}, \mathbf{n}) = \delta_{(\mathbf{0}, \mathbf{e}_z)}(\mathbf{y}, \mathbf{n})$. We denote this kernel with $p_t(\mathbf{y}, \mathbf{n})$. Our approach relies on the embedding in the Lie group $SE(3)$, on which the equation becomes:

$$\begin{cases} \partial_t \tilde{W}(\mathbf{y}, \mathbf{R}, t) = D_{33} \mathcal{A}_3^2 \tilde{W}(\mathbf{y}, \mathbf{R}, t) + D_{44} (\mathcal{A}_4^2 + \mathcal{A}_5^2) \tilde{W}(\mathbf{y}, \mathbf{R}, t), \\ \tilde{W}(\mathbf{y}, \mathbf{R}, 0) = \tilde{U}(\mathbf{y}, \mathbf{R}), \end{cases} \quad (4.2)$$

with \mathcal{A}_i the left-invariant vector fields, and $\tilde{W} : SE(3) \times \mathbb{R} \rightarrow \mathbb{R}$. The kernel on the group is denoted by $\tilde{p}_t(\mathbf{y}, \mathbf{R})$, subscript t indicating the evolution time. We use again a \sim to distinguish operators and functions on the group from operators and functions on the quotient $\mathbb{R}^3 \rtimes S^2$.

The advantage of having a closed formula for the kernel is that pre-computation allows for efficient implementation of the diffusion of dMRI images [RDtHRV10]. Recently, these kernels have been implemented by Stephan Meesters [MSG⁺16a] in the Diffusion Imaging in Python (Dipy) package [GBA⁺14], for contextual enhancement purposes. The effect of such enhancements on tractography methods in dMRI are treated in Chapter 6. For a broader embedding of the topic of hypo-elliptic operators in the literature, please see the introduction of Chapter 3.

4.1.1 Contributions and outline

The chapter is organized as follows. In Section 4.2, we further discuss the relation between the kernel on $SE(3)$ and on the group quotient $\mathbb{R}^3 \rtimes S^2$, and derive two symmetry/invariance properties, that we take as a requirement for any approximation. In Section 4.3 we show how the group $SE(3)$ can be approximated with a Heisenberg-type nilpotent group, and define a norm on this group that is equivalent (in the sense of norms) with the sub-Riemannian distance on $SE(3)$. We use this norm in Section 4.4 to give an analytic approximation of the diffusion kernel, show that it satisfies the right invariances, and that it gives a good approximation for reasonable parameter settings. We compare the approximation with the numerical solution of the previous chapter, that was derived by discretization and truncation of the series expression for the exact solution (for shortness abusively referred to as the ‘exact solution’). The stochastic approach is described in more detail in Section 4.5.

4.2 Symmetries of the diffusion kernel

Recall the definition of the group quotient $\mathbb{R}^3 \rtimes S^2$ in 2.1.1 and Definition 2.2.1 for legal operators. We define the \mathbb{L}_2 -space of functions on $SE(3)$ that can be identified with functions on the quotient $\mathbb{R}^3 \rtimes S^2$ as:

$$\mathbb{L}_2^R(SE(3)) := \{\tilde{U} \in \mathbb{L}_2(SE(3)) \mid \tilde{U}(gh) = \tilde{U}(g) \text{ for all } g \in SE(3), h \in H\},$$

with $H = \{(\mathbf{0}, \mathbf{R}) \mid \mathbf{R} = \mathbf{R}_{\mathbf{e}_z, \alpha}\}$. Here $\mathbf{R}_{\mathbf{n}, \phi}$ denotes a counter-clockwise rotation of angle ϕ about axis \mathbf{n} , later we also use $\mathbf{R}_{\mathbf{n}}$ to denote any rotation matrix that maps \mathbf{e}_z onto \mathbf{n} . Let $\tilde{\Phi} : \mathbb{L}_2^R(SE(3)) \rightarrow \mathbb{L}_2^R(SE(3))$ be linear and legal, and assume it maps $\mathbb{L}_2(SE(3))$ into $\mathbb{L}_\infty(SE(3))$. The fact that $\tilde{\Phi}$ is legal implies that there exists a $\Phi : \mathbb{L}_2(\mathbb{R}^3 \rtimes S^2) \rightarrow \mathbb{L}_2(\mathbb{R}^3 \rtimes S^2)$, such that $U \leftrightarrow \tilde{U} \implies \Phi(U) \leftrightarrow \tilde{\Phi}(\tilde{U})$, in the sense of (2.5). To ensure that any approximation of the diffusion kernel on the group is also well-defined on the quotient $\mathbb{R}^3 \rtimes S^2$, we need to take care of certain symmetries, as given in the following lemma and corollary.

Lemma 4.2.1. *There exists $\tilde{k} : SE(3) \times SE(3) \rightarrow \mathbb{R}$ and $k : \mathbb{R}^3 \rtimes S^2 \times \mathbb{R}^3 \rtimes S^2 \rightarrow \mathbb{R}$, such that for all $g = (\mathbf{y}, \mathbf{R}_{\mathbf{n}}) \in SE(3)$ we have*

$$(\tilde{\Phi}(\tilde{U}))(g) = \int_{SE(3)} \tilde{k}(g, q) \tilde{U}(q) dq = \int_{\mathbb{R}^3 \rtimes S^2} k((\mathbf{y}, \mathbf{n}), (\mathbf{y}', \mathbf{n}')) U((\mathbf{y}', \mathbf{n}')) d\mathbf{y}' d\sigma(\mathbf{n}') = (\Phi(U))(\mathbf{y}, \mathbf{n}).$$

Furthermore,

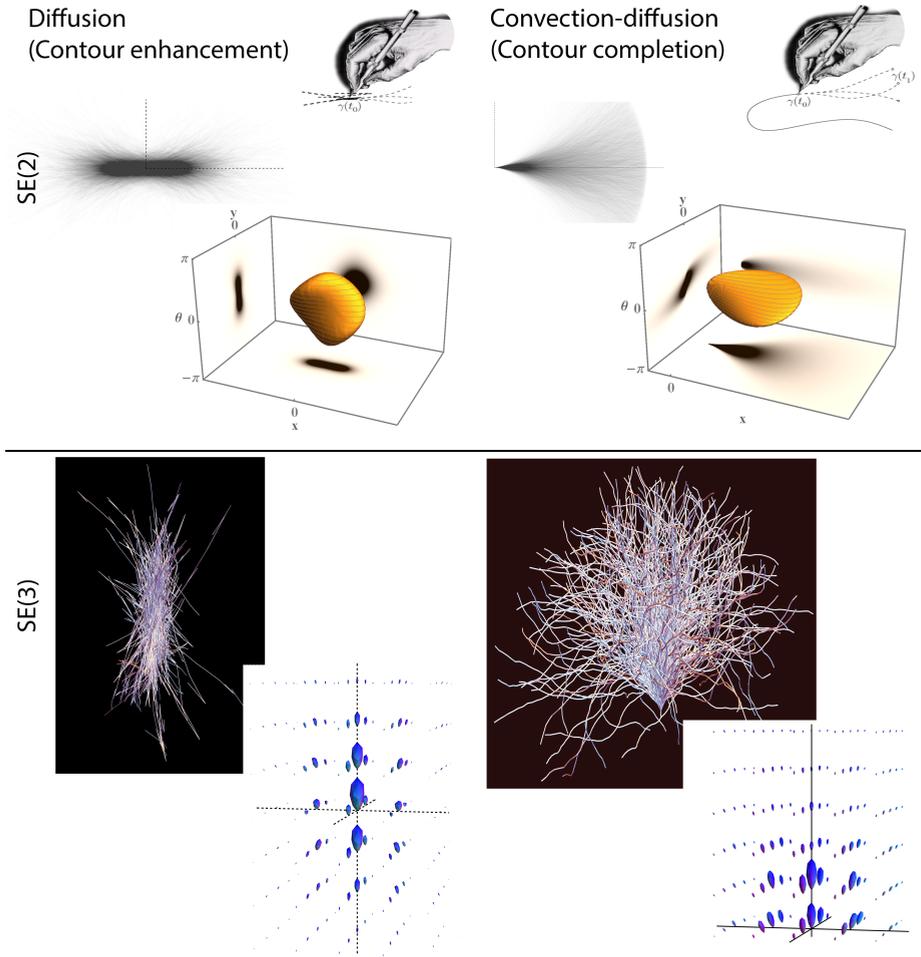


Figure 4.1: Various visualizations of the diffusion process (for contour enhancement, left), and the convection-diffusion process (for contour completion, right). For the SE(2) case, both the projections of random paths on \mathbb{R}^2 are shown, as well as an isocontour in SE(2) of the limiting distribution. For the SE(3) case, we show spatial \mathbb{R}^3 -projections of random paths in SE(3), and visualize the limiting distribution as a glyph field, as explained Section 1.1.1.

1. The kernel on the group satisfies $\tilde{k}(g', g) = \tilde{k}(g^{-1}g, e)$ and $\tilde{k}(g'h', gh) = \tilde{k}(g', g)$ for all $g, g' \in SE(3)$, $h, h' \in H$.
2. The kernel on the quotient satisfies $k((\mathbf{y}, \mathbf{n}), (\mathbf{y}', \mathbf{n}')) = k(\mathbf{R}_{\mathbf{n}'}^{-1}(\mathbf{y}-\mathbf{y}'), \mathbf{R}_{\mathbf{n}'}^{-1}\mathbf{n}), (\mathbf{0}, \mathbf{e}_z))$ and $k((\mathbf{R}_{\mathbf{e}_z, \alpha}\mathbf{y}, \mathbf{R}_{\mathbf{e}_z, \alpha}\mathbf{n}), (\mathbf{0}, \mathbf{e}_z)) = k((\mathbf{y}, \mathbf{n}), (\mathbf{0}, \mathbf{e}_z))$ for $(\mathbf{y}, \mathbf{n}), (\mathbf{y}', \mathbf{n}') \in \mathbb{R}^3 \times S^2$, $h \in H$.

Proof. It follows from the Dunford-Pettis Theorem that $\tilde{\Phi}$ and Φ are kernel operators. We can write

$$\begin{aligned} (\tilde{\Phi}(\tilde{U}))(\mathbf{y}, \mathbf{R}_n) &= \int_{SE(3)} \tilde{k}((\mathbf{y}, \mathbf{R}_n), (\mathbf{y}', \mathbf{R}')) \tilde{U}(\mathbf{y}', \mathbf{R}') d\mathbf{y}' d\mathbf{R}' \\ &= 2\pi \int_{\mathbb{R}^3 \times S^2} \tilde{k}((\mathbf{y}, \mathbf{R}_n), (\mathbf{y}', \mathbf{R}_{n'})) \tilde{U}(\mathbf{y}', \mathbf{R}_{n'}) d\mathbf{y}' d\sigma(\mathbf{n}') = \Phi(U)(\mathbf{y}, \mathbf{n}), \end{aligned} \quad (4.3)$$

which implies that $2\pi\tilde{k}((\mathbf{y}, \mathbf{R}_n), (\mathbf{y}', \mathbf{R}_{n'})) = k((\mathbf{y}, \mathbf{n}), (\mathbf{y}', \mathbf{n}'))$.

The identities in 1. follow immediately from the two properties of legal operators, for example by writing them in integral form. The identities in 2. follow from 1. together with the just established relation between \tilde{k} and k . \square

This lemma, as we will see in the next corollary, induces a type of invariance with respect to rotations around the reference axis. Another type of symmetry comes from the fact that the diffusion equation (4.2) is symmetric in forward and backward spatial directions. In other words, it has a reflectional invariance $(\mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5) \mapsto (-\mathcal{A}_3, -\mathcal{A}_4, -\mathcal{A}_5)$ in the diffusion. This yields the following corollary:

Corollary 4.2.2. *We can define kernels $\tilde{p}_t : SE(3) \rightarrow \mathbb{R}^+$, $\tilde{p}(g) := \tilde{k}(g, e)$ and $p_t : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}^+$, $p(\mathbf{y}, \mathbf{n}) := k((\mathbf{y}, \mathbf{n}), (\mathbf{0}, \mathbf{e}_z))$, and they satisfy the following symmetries*

$$p_t(\mathbf{y}, \mathbf{n}) = p_t(\mathbf{R}_{\mathbf{e}_z, \alpha} \mathbf{y}, \mathbf{R}_{\mathbf{e}_z, \alpha} \mathbf{n}) \text{ and } \tilde{p}_t(h^{-1}g) = \tilde{p}_t(g) = \tilde{p}_t(gh') \quad (4.4)$$

for all $t > 0$, $\alpha > 0$, $(\mathbf{y}, \mathbf{n}) \in \mathbb{R}^3 \times S^2$, $g \in SE(3)$, $h, h' \in H$. Moreover,

$$p_t(\mathbf{y}, \mathbf{n}) = p_t(-\mathbf{R}_n^T \mathbf{y}, \mathbf{R}_n^T \mathbf{e}_z) \text{ and } \tilde{p}_t(g) = \tilde{p}_t(g^{-1}). \quad (4.5)$$

Proof. The symmetry (4.4) is due to Lemma 4.2.1. The second symmetry (4.5) is due to reflectional invariance $(\mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5) \mapsto (-\mathcal{A}_3, -\mathcal{A}_4, -\mathcal{A}_5)$ in the diffusion (4.2) and reflection on the Lie algebra corresponds to inversion on the group. \square

Remark 4.2.3. *Note that (4.5) in terms of k would be: $k((\mathbf{y}, \mathbf{n}), (\mathbf{0}, \mathbf{e}_z)) = k((\mathbf{0}, \mathbf{e}_z), (\mathbf{y}, \mathbf{n}))$, by the relation $k((\mathbf{y}, \mathbf{n}), (\mathbf{y}', \mathbf{n}')) = p(\mathbf{R}_n^T(\mathbf{y} - \mathbf{y}'), \mathbf{R}_n^T \mathbf{n})$. This means that k defines a symmetric measure: evaluation in (\mathbf{y}, \mathbf{n}) of a kernel centered around the unity element $(\mathbf{0}, \mathbf{e}_z)$ should be equal to evaluation in the unity element of a kernel centered around (\mathbf{y}, \mathbf{n}) .*

Our goal of the next section will be to derive an approximation of sub-Riemannian distances, that we use later to give an approximation of the heat kernel that has the symmetries we have just derived.

4.3 Nilpotent approximation on $SE(3)$ and sub-Riemannian distances

The approximation of the Lie group $SE(3)$ by a nilpotent Lie group can be used to obtain an approximation of the sub-Riemannian distances on $SE(3)$. In addition to the Lie group theory discussed in Section 2.1, we need here the exponential and logarithmic map and the Lie bracket.

Exponential and logarithmic map We write $\mathfrak{se}(3)$ for the Lie algebra of $SE(3)$, which corresponds to the tangent space at the identity $g = (\mathbf{0}, \text{Id})$. The exponential map $\text{Exp} : \mathfrak{se}(3) \rightarrow SE(3)$ defines a mapping from a vector $X \in \mathfrak{se}(3)$ to an element in the group $SE(3)$. This element is found by following an integral curve along the left-invariant vector field $(L_g)_*X$. Here $(L_g)_*$ is the push-forward of the left-multiplication. The logarithmic map $\text{Log} : SE(3) \rightarrow \mathfrak{se}(3)$ is, as usual, the inverse of the exponential map.

We can express each element in the Lie algebra $\mathfrak{se}(3)$ in terms of the basis given in (2.7), with coefficients $\mathbf{c} = (c^1, \dots, c^6)^T$. Furthermore, we define $\mathbf{c}^{(1)} := (c^1, c^2, c^3)^T$ and $\mathbf{c}^{(2)} := (c^4, c^5, c^6)^T$, the spatial and rotational coefficients, respectively. We can make the exponential map $\text{Exp}_{SE(3)} : \mathfrak{se}(3) \rightarrow SE(3)$ and logarithmic map $\text{Log}_{SE(3)} : SE(3) \rightarrow \mathfrak{se}(3)$ explicit using these coefficients. The coefficients c^4, c^5, c^6 as in the 3×3 matrix Ω of the form

$$\Omega := \begin{pmatrix} 0 & -c^6 & c^5 \\ c^6 & 0 & -c^4 \\ -c^5 & c^4 & 0 \end{pmatrix}, \quad (4.6)$$

relate to the rotation $\mathbb{R} \in SO(3)$ via the matrix exponential, i.e., $\mathbf{R} = \exp(\Omega)$. The relation between the spatial coefficients $\mathbf{c}^{(1)}$ and (\mathbf{x}, \mathbf{R}) is given by

$$\mathbf{c}^{(1)} = \left(I - \frac{1}{2}\Omega + q^{-2} \left(1 - \frac{q}{2} \cot\left(\frac{q}{2}\right) \right) \Omega^2 \right) \mathbf{x}, \quad (4.7)$$

where $q = \|\mathbf{c}^{(2)}\|$ and Ω . Now

$$\text{Log}_{SE(3)}(g) = \sum_{i=1}^6 c_i(g) A_i, \quad \text{and} \quad \text{Exp}_{SE(3)} \left(\sum_{i=1}^6 c_i(g) A_i \right) = g, \quad (4.8)$$

using the relations above.

The Lie bracket and the BCH formula The Lie bracket for vector fields is defined as follows

$$[X, Y] := \lim_{t \rightarrow 0} \frac{\gamma(t) - e}{t^2}, \quad \text{with} \quad (4.9)$$

$$\gamma(t) = \text{Exp}(-tY) \text{Exp}(-tX) \text{Exp}(tY) \text{Exp}(tX).$$

It describes the infinitesimal displacement that occurs from following a path forth and back in X and Y directions. (The Lie bracket of two vector fields results again in a vector field.) The matrix of all commutators of the basis of the Lie algebra $\mathfrak{se}(3)$ is given by

$$[A_i, A_j]_{i,j=1}^6 = \begin{pmatrix} 0 & 0 & 0 & 0 & A_3 & -A_2 \\ 0 & 0 & 0 & -A_3 & 0 & A_1 \\ 0 & 0 & 0 & A_2 & -A_1 & 0 \\ 0 & A_3 & -A_2 & 0 & A_6 & -A_5 \\ -A_3 & 0 & A_1 & -A_6 & 0 & A_4 \\ A_2 & -A_1 & 0 & A_5 & -A_4 & 0 \end{pmatrix} \quad (4.10)$$

For two left-invariant vector fields $X = \sum_{i=1}^6 x^i \mathcal{A}_i$ and $Y = \sum_{i=1}^6 y^i \mathcal{A}_i$ the Baker-Campbell-Hausdorff (BCH) formula (see e.g. [Ros06]) gives:

$$\text{Log}(\text{Exp}(X) \text{Exp}(Y)) = X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}([X, [X, Y]] + [Y, [Y, X]]) + \mathcal{O}([\cdot, [\cdot, [\cdot, \cdot]]]), \quad (4.11)$$

where $\mathcal{O}([\cdot, [\cdot, [\cdot, \cdot]]])$ denotes higher order nested brackets. Since the Lie algebra $\mathfrak{se}(3)$ is not nilpotent (it has non-vanishing Lie brackets of order ≥ 2), the BCH formula gives an infinite series of nested Lie brackets. We truncate the BCH expansion by omitting all Lie brackets of order 2 (once nested brackets) or higher, as if the Lie algebra $\mathfrak{se}(3)$ is nilpotent of step 2. This gives the approximation

$$\text{Log}(\text{Exp}(X) \text{Exp}(Y)) \approx X + Y + \frac{1}{2}[X, Y]. \quad (4.12)$$

Only the vector fields $\mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5$ appear in the generator of our diffusion, and their commutators correspond to the remaining vector fields $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_6$, as can be seen from the commutator relations in (4.10). Regarding $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_6$ already as commutators of the other three, they are also ignored in the approximation (4.12). We can now use this approximation to define a group product on the vector space \mathbb{R}^6 of canonical coordinates of the first kind:

$$(x^1, \dots, x^6) \cdot (y^1, \dots, y^6) = \begin{pmatrix} x^1 + y^1 + \frac{1}{2}(x^5 y^3 - x^3 y^5) \\ x^2 + y^2 + \frac{1}{2}(x^3 y^4 - x^4 y^3) \\ x^3 + y^3 \\ x^4 + y^4 \\ x^5 + y^5 \\ x^6 + y^6 + \frac{1}{2}(x^4 y^5 - x^5 y^4) \end{pmatrix}^T, \quad (4.13)$$

with x^i, y^i coordinates of the first kind given by the logarithmic map (4.6) and (4.7). This new group is a free-nilpotent group of rank 3 and step 2. Groups that are nilpotent allow for homogeneous norms on the group elements. Via a scaling of the spatial generators $\tilde{A}_i = \xi^{-1} A_i$, $i \in \{1, 2, 3\}$, to balance between the spatial and angular scale, we define the following ξ -isotropic norm on the coefficients:

$$\|\text{Log}_{SE(3)}(g)\|_{\xi, \zeta} := \|\mathbf{c}\|_{\xi, \zeta} := \sqrt[4]{(\xi^2 |c^3|^2 + |c^4|^2 + |c^5|^2)^2 + \zeta (\xi^2 (|c^1|^2 + |c^2|^2) + |c^6|^2)}, \quad (4.14)$$

where $\mathbf{c} = \mathbf{c}(g)$ according to (4.8), and $\zeta > 0$. Here we followed the approach of [DF11] by adapting the logarithmic approximation to a differentiable one, which explains the fourth root above.

Sub-Riemannian distances Instead of using such a norm to measure the distance between two elements, we could measure the distance between two elements in $SE(3)$ by considering a sub-Riemannian geometry and measuring the lengths of shortest horizontal paths. We define a distribution $\Delta := \text{span}\{\mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}$, a sub-bundle of the full tangent bundle $T(SE(3)) := \text{span}\{\mathcal{A}_i\}_{i=1}^6$. This distribution is chosen to have the same vector fields as the one generating the diffusion. A horizontal path is then a curve $\gamma : [t_0, t_1] \subset \mathbb{R} \rightarrow SE(3)$, such that its tangent lies in the distribution for all $t \in [t_0, t_1]$. They can be written as

$$\dot{\gamma}(t) = u^3(t) \mathcal{A}_3|_{\gamma(t)} + u^4(t) \mathcal{A}_4|_{\gamma(t)} + u^5(t) \mathcal{A}_5|_{\gamma(t)},$$

where u^3 is a spatial control, and u^4 and u^5 are angular controls of the curve. Their length can be measured using the sub-Riemannian metric tensor

$$\mathcal{G}^{\xi, \zeta}|_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t)) := \mathcal{C}(\gamma(t))^2 (\xi^2 |u^3(t)|^2 + (|u^4(t)|^2 + |u^5(t)|^2)), \quad (4.15)$$

in which $\mathcal{C} : SE(3) \rightarrow \mathbb{R}^+$ is an external cost which penalizes the curves to move through certain regions in $SE(3)$, ξ is the same parameter as in (4.14), which balances the penalty of motion in the angular and spatial directions and has dimensions [1/length].

The sub-Riemannian distance between two points $g_1, g_2 \in SE(3)$ is then given by

$$d_0(g_1, g_2) := \inf \left\{ \int_0^1 \sqrt{\mathcal{G}^{\xi, \mathcal{C}}|_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))} dt \right\}, \quad (4.16)$$

where the infimum is taken over Lipschitz continuous curves $\gamma \in \text{Lip}([0, T], SE(3))$ with $\gamma(0) = g_1$, $\gamma(1) = g_2$, and $\dot{\gamma}(t) = \Delta|_{\gamma(t)}$. Note that due to the inclusion of an external cost function \mathcal{C} the distance d is not strictly left-invariant, however, when substituting \mathcal{C} by $\mathcal{C}_g := \mathcal{C}(g^{-1}h)$ in (4.16) we do have left-invariance (i.e., then $d(g \cdot g_1, g \cdot g_2) = d(g_1, g_2)$). In the remainder of this chapter, we take $\mathcal{C} = 1$, in Chapters 5 and 7 the same problem is discussed for a data-dependent cost \mathcal{C} .

4.3.1 Approximation of the sub-Riemannian distance

Finally we arrive at the sub-Riemannian distance approximations. By the Ball-Box theorem (see e.g. [Bel96]) and equivalence of homogeneous norms, there exists a constant \mathbf{c} such that

$$\mathbf{c}^{-1} \|\text{Log}(g)\|_{\zeta} \leq d_0(e, g) \leq \mathbf{c} \|\text{Log}(g)\|_{\zeta},$$

with $\text{Log}(g)$ defined by Eqs. (4.6) and (4.7). The logarithmic norm is locally equivalent to the sub-Riemannian distance, which was proved in full generality in [NSW85, Thm. 2 & 4], covering also the current setting, and further used in [JS86] to provide heat kernel estimates.

With numerical experiments we find that the norm $\|\cdot\|_{\xi, \zeta}$ closely approximates the sub-Riemannian distance for $\mathcal{C} = 1$ via

$$d_0(g, h) \approx \|\text{Log}(g^{-1}h)\|_{\xi, \zeta} = \|\mathbf{c}\|_{\xi, \zeta}, \quad (4.17)$$

with \mathbf{c} the coordinates of the first kind obtained via Eqs. (4.6), (4.7).

For $\zeta = 16$ this norm coincides with the well-known Folland-Kaplan-Korányi gauge, which is a widely studied norm on Carnot groups due to its relation to fundamental solutions of sub-Laplacians [BLU07]: Folland found that $\|\mathbf{c}\|_{16}^{2-Q}$, with homogeneous dimensions Q , is (a constant multiple of) the fundamental solution of the canonical sub-Laplacian on the Heisenberg group [Fol73]; Kaplan showed that this relation more generally holds for H-type (Carnot) groups [Kap80]; Korányi derived many more of its properties in relation to harmonic analysis and potential theory [Kor82].

In view of the Folland-Kaplan-Korányi gauge setting $\zeta = 16$ in $\|\cdot\|_{\xi, \zeta}$ would be a sensible choice. However, the visual comparison in Fig. 4.2 indicates that better approximations could be obtained for $\zeta = 100$. For a quantitative comparison between the sub-Riemannian distance and the logarithmic norm via the nilpotent approximation (for both $SE(2)$ and $SE(3)$, see [BCP17, Appendix A]). Nevertheless, setting the parameter

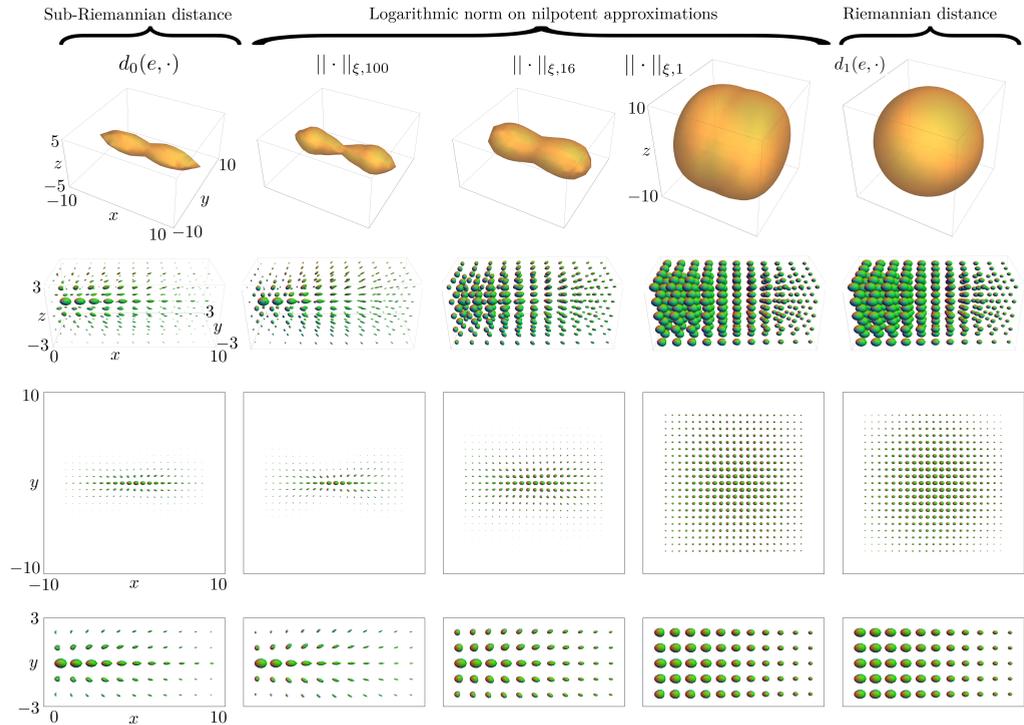


Figure 4.2: Distances on $SE(3)$ for $\xi = .1$, $C = 1$, with the origin placed at $e = (\mathbf{0}, \mathbf{e}_x)$. Top row: Level sets of the spatial projections (minimum intensity projections over S^2) of the distance volumes on $SE(3)$. Rows two to four: Glyph visualizations in which each distance volume d is visualized with a "Gaussian" density $U(g) = e^{-d(e,g)^2}$. For an interpretation of the glyphs see Section 1.1.1. Row two: Glyph visualizations of a sub-volume. Row three: Glyph visualization of a slice at $z = 0$. Row four: Zoomed in glyph visualization of a slice at $z = 0$. From left to right: The sub-Riemannian distance $d_0(e, \cdot)$ on $SE(3)$, see Eq. (4.16); Homogeneous norms $\|\cdot\|_{\xi, \zeta}$, see Eq. (4.14), of the nilpotent approximation $(SE(3))_0$ for respectively $\zeta = 100$, $\zeta = 16$ (Folland-Kaplan-Korányi gauge) and $\zeta = 1$; The (ξ -isotropic) Riemannian distance $d_1(e, \cdot)$ on $SE(3)$.

$\zeta = 16$ does seem appropriate in the next section, where we use the norm to approximate the heat kernel.

4.4 New vs. previous kernel approximations

In [RDtHRV10] a possible approximation kernel for p_t is presented, based on a direct product of two $SE(2)$ -kernels. We will refer to this function as $p_t^{\text{prev},1}$. This approximation is easy to use since the value $p_t^{\text{prev},1}(\mathbf{y}, \mathbf{n})$ is defined directly in terms of the spherical coordinates of \mathbf{n} . However, in Fig. 4.3 we show that the symmetries described before are

not preserved by $p_t^{\text{prev},1}$ and errors tend to be larger when D_{44} and t increase. To avoid this problem, we make use of the norm on the logarithm of the previous section, that, when treated correctly, does satisfy the appropriate symmetry.

The approximation is based on the known observation, see e.g. [BBN12], that

$$\tilde{p}_t(g) = Ce^{-\frac{d_0(\epsilon, g)^2}{4t}}, \quad (4.18)$$

where the constant C can depend on t . An approximation of the kernel that makes use of the norm $\|\cdot\|_{\xi, \zeta}$ of the sub-Riemannian distance was given by [DF11]:

$$\tilde{p}_t^{\log}(\mathbf{c}(g)) := Ce^{-\frac{\|\log_{SE(3)}(g)\|_{\xi, \zeta}^2}{4t}}, \quad (4.19)$$

with $C = (4\pi t^2 D_{33} D_{44})^{-2}$, $\xi = \sqrt{D_{44}/D_{33}}$, and $\zeta = 1$. Now the difficulty lies in the fact that the logarithm is only well-defined on the group $SE(3)$, not on the quotient $\mathbb{R}^3 \times S^2$, and explicitly depends on the choice of α in the rotational part $\mathbf{R}_{\gamma, \beta, \alpha} := \mathbf{R}_{\mathbf{e}_z, \gamma} \mathbf{R}_{\mathbf{e}_y, \beta} \mathbf{R}_{\mathbf{e}_z, \alpha}$ of the group element. It is therefore not straightforward to use this approximation kernel such that the invariance properties in Corollary 4.2.2 are preserved. In previous work the choice of $\alpha = 0$ is taken, giving rise to the approximation

$$p_t^{\text{prev},2}(\mathbf{y}, \mathbf{n}(\beta, \gamma)) = \tilde{p}_t^{\log}(\mathbf{c}(\mathbf{y}, \mathbf{R}_{\mathbf{e}_z, \gamma} \mathbf{R}_{\mathbf{e}_y, \beta})). \quad (4.20)$$

However, this section does not satisfy the appropriate symmetries, as is pointed out in Fig. 4.3. In contrast, we propose to take the section $\alpha = -\gamma$, which indeed provides us the correct symmetry for the Gaussian approximation of $p_t(\mathbf{y}, \mathbf{n})$. Two key ingredients for this are: (1) with this choice $c^6 = 0$, recall 4.6, and (2) suppose \mathbf{n} and $\mathbf{R}_{\mathbf{n}}^T \mathbf{e}_z = \mathbf{m}$ are given, then:

$$\mathbf{R}_{\mathbf{n}} = \mathbf{R}_{\mathbf{e}_z, \gamma} \mathbf{R}_{\mathbf{e}_y, \beta} \mathbf{R}_{\mathbf{e}_z, \alpha} \quad \text{and} \quad \mathbf{R}_{\mathbf{m}} = \mathbf{R}_{\mathbf{e}_z, -\alpha} \mathbf{R}_{\mathbf{e}_y, -\beta} \mathbf{R}_{\mathbf{e}_z, \alpha'},$$

with α, α' free. Only for the choice $\alpha = -\gamma$, we see that we $\mathbf{R}_{\mathbf{n}}^T = \mathbf{R}_{\mathbf{m}}$, which yields the right invariance, as stated in the following theorem.

Theorem 4.4.1. *When the approximate kernel p_t^{new} on the quotient is related to the approximate kernel on the group \tilde{p}_t^{\log} by*

$$p_t^{\text{new}}(\mathbf{y}, \mathbf{n}(\beta, \gamma)) := \tilde{p}_t^{\log}(\mathbf{c}(\mathbf{y}, \mathbf{R}_{\mathbf{e}_z, \gamma} \mathbf{R}_{\mathbf{e}_y, \beta} \mathbf{R}_{\mathbf{e}_z, -\gamma})), \quad (4.21)$$

i.e. we make the choice $\alpha = -\gamma$, we have the desired α -left-invariant property

$$p_t^{\text{new}}(\mathbf{y}, \mathbf{n}) = p_t^{\text{new}}(\mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{y}, \mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{n}), \quad \alpha' \in [0, 2\pi]. \quad (4.22)$$

and the symmetry property

$$p_t^{\text{new}}(\mathbf{y}, \mathbf{n}) = p_t^{\text{new}}(-\mathbf{R}_{\mathbf{n}}^T \mathbf{y}, \mathbf{R}_{\mathbf{n}}^T \mathbf{e}_z) \quad (4.23)$$

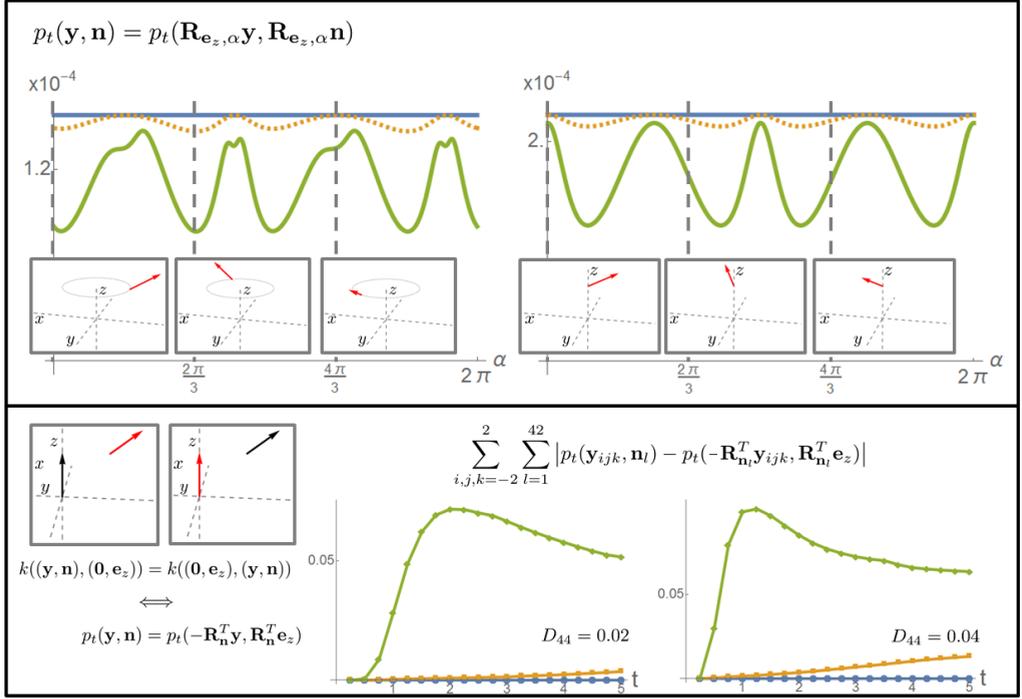


Figure 4.3: *Top*: 2 arbitrary positions and orientations, rotated around the z -axis. The graphs display $p_t(\mathbf{R}_{\mathbf{e}_z, \alpha} \mathbf{y}, \mathbf{R}_{\mathbf{e}_z, \alpha} \mathbf{n})$, for $p_t^{\text{prev},1}$ in green, $p_t^{\text{prev},2}$ in orange and p_t^{new} in blue. Parameters $D_{33} = 1, D_{44} = 0.02, t = 4$. *Bottom*: the equality $p_t(\mathbf{y}, \mathbf{n}) = p_t(-\mathbf{R}_{\mathbf{n}}^T \mathbf{y}, \mathbf{R}_{\mathbf{n}}^T \mathbf{e}_z)$ should hold. We show the error as indicated in the figure, with $\mathbf{y}_{ijk} = (i, j, k)^T$ and \mathbf{n}_l uniformly distributed over S^2 , for $p_t^{\text{prev},1}$, $p_t^{\text{prev},2}$ and p_t^{new} . Parameter $D_{33} = 1$ and t, D_{44} are as shown.

Proof. We start by proving the α -invariance. Following definition (4.21) we have

$$p_t^{\text{new}}(\mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{y}, \mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{n}(\beta, \gamma)) = \tilde{p}_t^{\text{log}}(\mathbf{c}(\mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{y}, \mathbf{R}_{\gamma + \alpha'} \mathbf{R}_{\beta} \mathbf{R}_{-(\gamma + \alpha')}))$$

We define $\Omega_{\gamma, \beta, \alpha}$ by $\mathbf{R}_{\gamma, \beta, \alpha} =: e^{\Omega_{\gamma, \beta, \alpha}}$. Therefore in our case, where we choose $\alpha = -\gamma$, we find for all α' :

$$\begin{aligned} \mathbf{R}_{\mathbf{e}_y, \beta} &= \mathbf{R}_{\mathbf{e}_z, \gamma + \alpha'}^{-1} e^{\Omega_{\gamma + \alpha', \beta, -(\gamma + \alpha')}} \mathbf{R}_{\mathbf{e}_z, \gamma + \alpha'} = e^{\mathbf{R}_{\mathbf{e}_z, \gamma + \alpha'}^{-1} (\Omega_{\gamma + \alpha', \beta, -(\gamma + \alpha')}) \mathbf{R}_{\mathbf{e}_z, \gamma + \alpha'}} \\ &= e^{\mathbf{R}_{\mathbf{e}_z, \gamma}^{-1} \mathbf{R}_{\mathbf{e}_z, \alpha'}^{-1} (\Omega_{\gamma + \alpha', \beta, -(\gamma + \alpha')}) \mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{R}_{\mathbf{e}_z, \gamma}} \end{aligned}$$

We see that $\mathbf{R}_{\mathbf{e}_z, \alpha'} (\Omega_{\gamma, \beta, -\gamma}) \mathbf{R}_{\mathbf{e}_z, \alpha'}^{-1} = \Omega_{\gamma + \alpha', \beta, -(\gamma + \alpha')}$. From this we deduce:

$$\mathbf{c}_{\gamma + \alpha', \beta, -(\gamma + \alpha')}^{(2)} = \mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{c}_{\gamma, \beta, -\gamma}^{(2)}, \quad \text{and } q_{\gamma + \alpha', \beta, -(\gamma + \alpha')} = q_{\gamma, \beta, -\gamma} =: q, \quad (4.24)$$

and together with (4.7) it gives $\mathbf{c}_{\mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{x}, \gamma + \alpha', \beta, -(\gamma + \alpha')}^{(1)} = \mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{c}_{\mathbf{x}, \gamma, \beta, -\gamma}^{(1)}$. Combining

this with (4.24) gives

$$\mathbf{c}_{\mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{x}, \gamma + \alpha', \beta, -(\gamma + \alpha')} = Z_{\alpha'}^T \mathbf{c}_{\mathbf{x}, \gamma, \beta, -\gamma}, \quad \text{with } Z_{\alpha'} = \begin{pmatrix} \mathbf{R}_{\mathbf{e}_z, \alpha'}^T & 0 \\ 0 & \mathbf{R}_{\mathbf{e}_z, \alpha'}^T \end{pmatrix} \quad (4.25)$$

It follows immediately that $(c^1)^2 + (c^2)^2$, c^3 , $(c^4)^2 + (c^5)^2$ and c^6 are independent of α' . The proof for α -invariance is completed by stating that given $\alpha' \in [0, 2\pi]$:

$$\begin{aligned} p_t^{\text{new}}(\mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{y}, \mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{n}(\beta, \gamma)) &= \tilde{p}_t^{\log}(\log_{SE(3)}(\mathbf{R}_{\mathbf{e}_z, \alpha'} \mathbf{y}, \mathbf{R}_{\gamma + \alpha', \beta, -(\gamma + \alpha')})) = \\ \tilde{p}_t^{\log}(Z_{\alpha'}^T \log_{SE(3)}(\mathbf{y}, \mathbf{R}_{\gamma, \beta, -\gamma})) &= \tilde{p}_t^{\log}(\log_{SE(3)}(\mathbf{y}, \mathbf{R}_{\gamma, \beta, -\gamma})) = p_t^{\text{new}}(\mathbf{y}, \mathbf{n}(\beta, \gamma)). \end{aligned}$$

The symmetry property (4.23) directly follows from the fact that

$$g^{-1} = \exp_{SE(3)} \left(- \sum_{i=1}^6 c^i A_i \right) = (\mathbf{y}, \mathbf{R})^{-1} = (-\mathbf{R}^{-1} \mathbf{y}, \mathbf{R}^{-1}),$$

the fact that $\|\mathbf{c}\|_{\xi, \zeta}$ is invariant under reflection $\mathbf{c} \mapsto (-\mathbf{c})$, and the fact that the section $\alpha = -\gamma$ is invariant under inversion. \square

The visualization in Fig. 4.3 can be observed to be in line with what we have just shown, since the blue line shows no errors when checking for these two invariances. The right column in Fig. 4.4 shows a visualization comparing the numerical approximation with the approximation of Chapter 3, that was obtained by discretizing and truncating the series expansion of the exact solution. The parameters of the diffusion are $D_{33} = 1.$, $D_{44} = 0.1$ and $t \in \{1., 2.\}$. However, when computing the approximating kernel p_t^{new} , it seems to be overestimating the diffusion time. The results shown are obtained with $\xi = \sqrt{D_{44}/D_{33}}$, $\zeta^{-1} = 16$ and $t \in C\{1., 2.\}$, with a heuristically chosen $C = 0.6$. For this choice of parameters, the approximation (right column) resembles the ‘exact solution’ (left column) quite well. The advantage of this approximation is that it is fast and can be parallelized, and when the norm for certain ξ, ζ is precomputed and stored, the kernel can easily be computed for any desired diffusion time.

4.5 Stochastic approximation of diffusion kernels

The PDE for hypo-elliptic diffusion corresponds to the Fokker-Planck equation or forward Kolmogorov equation describing the evolution of a probability density function related to a certain stochastic process. The solution kernels can thereby also be approximated with a Monte Carlo simulation, where the end points of random walks, see Fig. 4.1, are accumulated. We briefly discuss how this simulation can be done.

Let $t > 0$, $D_{33} > 0$, $D_{44} > 0$. We simulate M random walks, with N steps. We denote the position after k steps with position $\mathbf{y}_k \in \mathbb{R}^3$ and the orientation with $\mathbf{n}_k \in S^2$. For the diffusion case, each random walk is given by:

$$\begin{cases} \mathbf{y}_{k+1} = \mathbf{y}_k + \varepsilon_{k+1} \sqrt{\frac{12tD_{33}}{N}} \mathbf{n}_k, & \varepsilon_k \sim \text{UNIF}(-1, 1), \\ \mathbf{n}_{k+1} = \mathbf{R}_{\mathbf{n}_k} \mathbf{R}_{\mathbf{e}_z, \gamma_{k+1}} \mathbf{R}_{\mathbf{e}_y, \beta_{k+1}} \sqrt{\frac{12tD_{44}}{N}} \mathbf{R}_{\mathbf{n}_k}^T \mathbf{n}_k, & \gamma_k \sim \text{UNIF}(-\pi, \pi), \beta_k \sim \text{UNIF}(-1, 1). \end{cases} \quad (4.26)$$

Intuitively, every step in the random walk consists of a small uniformly distributed step in the current orientation and a small change in orientation.

The distribution is then computed simply by binning the endpoints of the random walks. We divide \mathbb{R}^3 into cubes c_{ijk} , $i, j, k \in \mathbb{Z}$, of size $\Delta s \times \Delta s \times \Delta s$:

$$c_{ijk} := \left[\left(i - \frac{1}{2}\right)\Delta s, \left(i + \frac{1}{2}\right)\Delta s \right] \times \left[\left(j - \frac{1}{2}\right)\Delta s, \left(j + \frac{1}{2}\right)\Delta s \right] \times \left[\left(k - \frac{1}{2}\right)\Delta s, \left(k + \frac{1}{2}\right)\Delta s \right].$$

We divide S^2 into bins B_l , $l = \{1, \dots, b\}$ for $b \in \mathbb{N}$, with surface area σ_{B_l} and maximal surface area σ_B . The number of random walks in a simulation with traveling time t that have their end point $\mathbf{y}_N \in c_{ijk}$ with their orientation $\mathbf{n}_N \in B_l$ is denoted with $\#_t^{ijkl}$. Furthermore, we define the indicator function

$$\mathbb{1}_{c_{ijk}, B_l}(\mathbf{y}, \mathbf{n}) := \begin{cases} 1 & \mathbf{y} \in c_{ijk}, \mathbf{n} \in B_l, \\ 0 & \text{otherwise.} \end{cases}$$

Without proof we state here that when the number of paths $M \rightarrow \infty$, the number of steps in each path $N \rightarrow \infty$ and the bin sizes tend to zero, the obtained distribution converges to the exact kernel:

$$\lim_{\Delta s, \sigma(B) \rightarrow 0} \lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} p_t^{\Delta s, \sigma_B, N, M}(\mathbf{y}, \mathbf{n}) = p_t(\mathbf{y}, \mathbf{n}), \quad (4.27)$$

with

$$p_t^{\Delta s, \sigma_B, N, M}(\mathbf{y}, \mathbf{n}) = \sum_{l=1}^b \sum_{i, j, k \in \mathbb{Z}} \mathbb{1}_{c_{i, j, k}, B_l}(\mathbf{y}, \mathbf{n}) \frac{\#_t^{ijkl}}{M(\Delta s)^3 \sigma_{B_l}}. \quad (4.28)$$

The middle column of Fig. 4.4 shows the result of this Monte Carlo simulation, again with $D_{33} = 1$, $D_{44} = 0.1$ and $t \in \{1, 2\}$. The kernel is computed with bin size $\Delta s = 0.25$, and with $b = 162$, obtained by refining an icosahedron three times. We take $M = 10^8$ paths, and $N = 20$ steps on each path. The noise is still visible in the kernels, especially in the level sets, but the overall shape of the kernel is close to the kernel on the left. The advantage of this method is that it is very easy to implement and easy to parallelize, but the convergence to the exact solution is slow.

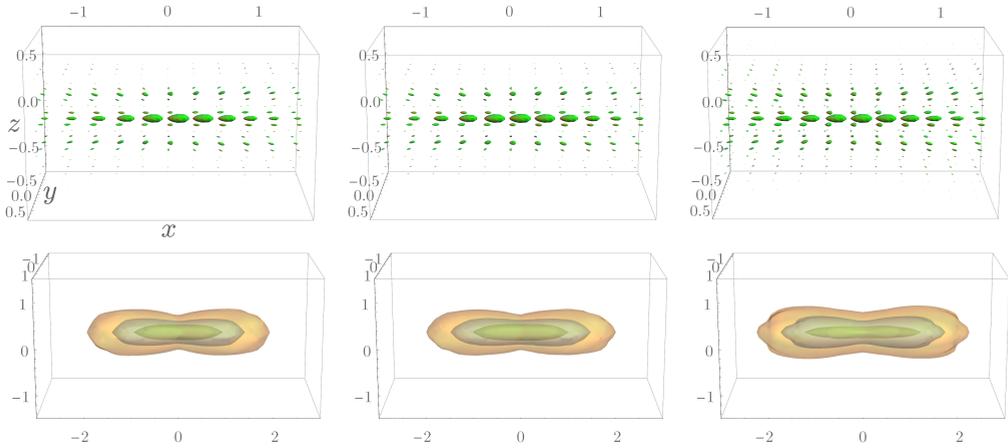
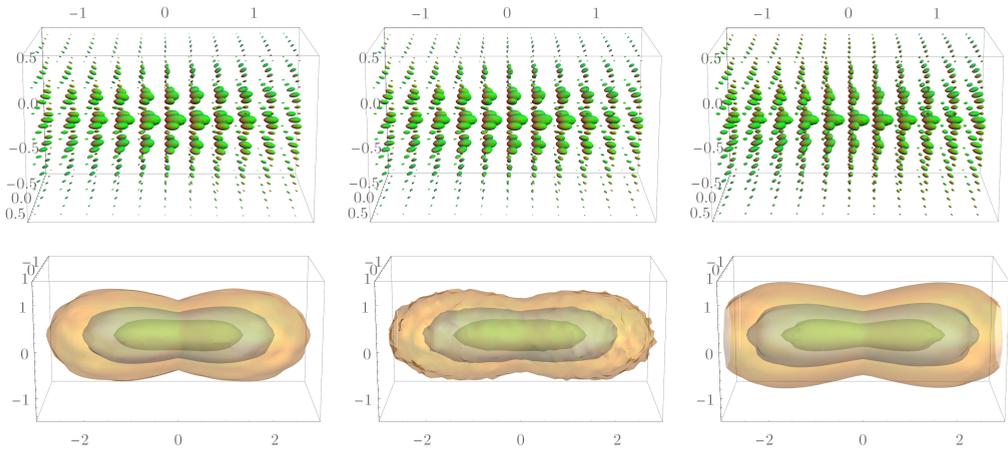
(a) Diffusion time $t = 1$.(b) Diffusion time $t = 2$.

Figure 4.4: Comparison of the kernels obtained with the exact approach (left), Monte Carlo simulation (middle) and the nilpotent approximation (right). The first and third row show glyph visualizations of the kernel, the second and fourth row visualize level sets for values 0.5 (yellow), 0.25 (blue), 0.1 (green). N.B. for easier visualization we displayed $p_t((0, \mathbf{R}_{e_x})^{-1}(\mathbf{y}, \mathbf{n}))$, i.e., the kernel is oriented along x -axis.

4.6 Conclusion

In this chapter, we have applied the known connections between heat equations, random walks and (sub-)Riemannian distances to show the correspondence between the heat kernels that can be obtained with these different approaches.

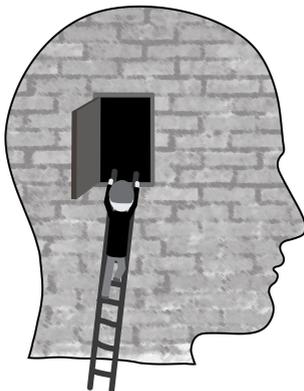
We have shown that the hypo-elliptic diffusion equation on position-orientation space imposes two invariance/symmetry properties on the heat kernel, that are ideally satisfied by any approximation kernel. However, we found that this was not the case for both approximations that existed so far.

Therefore, we worked out in detail how the theory of homogeneous norms can be applied by approximating the Lie group $SE(3)$ with a nilpotent group $(SE(3))_0$. The resulting norm, that approximates the sub-Riemannian distance on $SE(3)$ (without data-term), could be used to give an approximation kernel that does satisfy the right symmetries. Another approximation for the kernel could be obtained using a very simple and intuitive Monte Carlo simulation of random walks.

With the exact approaches of the previous chapter in hand, we found that, although noisy by nature, the Monte Carlo simulation provides a good approximation, see Fig. 4.4. Furthermore, it could be observed that the homogeneous norm on the nilpotent approximation $(SE(3))_0$ yields good approximation, albeit by choosing a lower diffusion time than for the exact solution. This would make it possible to improve the accuracy of fast implementations for solving the diffusion on dMRI images [RDtHRV10, MSG⁺16a].

Optimal Paths in $SE(2)$ and $SE(3)$

Based on: [DMMP16] R. Duits, S. P. L. Meesters, J-M. Mirebeau, and J. M. Portegies. Optimal Paths for Variants of the 2D and 3D Reeds-Shepp Car with Applications in Image Analysis. *arXiv:1612.06137 [math]*, December 2016. *Accepted for publication in JMIV, Special Issue ‘Differential Geometry and Orientation Analysis’.*



5.1 Introduction

Shortest paths in position and orientation space are central in this chapter. Dubins describes in [Dub57] the problem of finding shortest paths for a car in the plane between initial and final points and direction, with a penalization on the radius of curvature, for a car that has no reverse gear. Reeds and Shepp consider in [RS90] the same problem, but then for a car that does have the possibility for backward motion. In both papers, the focus lies on describing and proving the general shape of the optimal paths, without giving explicit solutions for the shortest paths.

This can be considered a curve optimization problem in the space $\mathbb{R}^2 \times (\mathbb{R}/2\pi\mathbb{Z})$, equipped with the natural Euclidean metric but only among curves $\gamma(t) = (x(t), y(t), \theta(t))$ subject to the constraint that $(\dot{x}(t), \dot{y}(t))$ is proportional to $(\cos \theta(t), \sin \theta(t))$. Formulating the problem this way, it becomes one of the simplest examples of sub-Riemannian (SR) geometry: the tangent vector $\dot{\gamma}(t)$ is constrained to remain in the span of $(\cos \theta(t), \sin \theta(t), 0)$ and $(0, 0, 1)$, see Fig. 5.1. The SR curve optimization problem and the properties of its geodesics in $\mathbb{R}^2 \times S^1$ have been studied and applied in image analysis by [Pet03, CS06, DBRS13, BDRS14, MAS13, AS04], and in particular for modelling the Reeds-Shepp car in [MS10, BCR10, Sac11], whereas the latter presented a complete and optimal synthesis for the geometric control problem on $\mathbb{R}^2 \times S^1$ with uniform cost. Properties of SR geodesics in $\mathbb{R}^d \times S^{d-1}$ with $d = 3$ have been studied in [DGDHM16] and for general d in [DGDHS14]. Apart from the Reeds-Shepp car problem, there are other examples relating optimal control theory and SR geometry, see for example the books by Agrachev and Sachkov [AS04] and Montgomery [Mon02]. Extensive results on the relation between sub-Riemannian distances on \mathbb{R}^n and viscosity solutions to Hamilton-Jacobi equations can be found in [Dra05, Dra07]. Applications in robotics and visual modeling of SR geometry and control theory can be found in e.g. [SBG⁺14].

On the left in Fig. 5.2, we show an example of an optimal path between two points in $\mathbb{R}^2 \times S^1$. The projection on \mathbb{R}^2 of this curve has two parts where the car moves in reverse (the red parts of the line), resulting in two cusps. From the perspective of image analysis applications this is undesirable and it is a valid question what the optimal paths are if cusps and reverse gear are not allowed. In this chapter, similar to the difference between the Dubins car and the Reeds-Shepp car, we also consider this variant: it can be accounted for by requiring that the spatial propagation is forward. This variant falls outside the SR framework and requires asymmetric Finsler geometry instead.

Furthermore, we would like to extend the Finsler metric using two data-driven factors that can vary with position and orientation. This can be used to compute shortest paths for a car, where for example road conditions and obstacles are taken into account. In [BDMS15] it is shown this approach is useful for tracking vessels in retinal images.

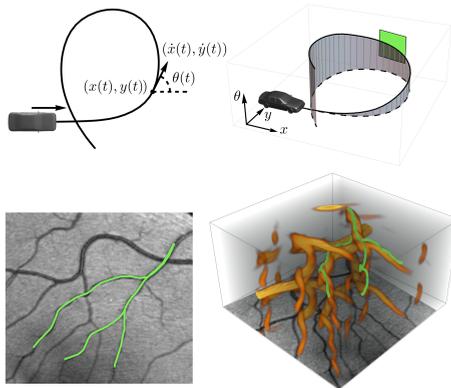


Figure 5.1: Top: A car can only move in its current orientation or change its current orientation. In other words, when the path $\gamma(t) = (x(t), y(t), \theta(t))$ is considered as indicated in the left figure, the tangent $\dot{\gamma}(t)$ is restricted to the span of $(\cos \theta(t), \sin \theta(t), 0)$ and $(0, 0, 1)$, of which the green plane on the right is an example. Bottom: the meaning of shortest path between points in an image is determined by a combination of a cost computed from the data, the restriction above, and a curvature penalization. The path optimization problem is formulated on the position-orientation domain such as in the image on the right. The cost for moving through the orange parts is lower than elsewhere.

Likewise, the 3D variant of the problem provides a basis for algorithms for blood vessel detection in 3D Magnetic Resonance Angiography (MRA) data, or detection of shortest paths and quantification of structural connectivity in 5D dMRI data of the brain. The paper [DMMP16] on which this chapter is based originally contained a validation of the methods on several applications. That section is extended for this thesis and therefore postponed to Chapter 7.

5.1.1 A distance function and shortest paths on $\mathbb{R}^d \times S^{d-1}$

We briefly repeat some definitions and notation from Section 2.3. With $d \in \{2, 3\}$, we define $\mathbb{M} := \mathbb{R}^d \times S^{d-1}$ to be the $2d-1$ dimensional manifold of positions and orientations. We use a Finsler metric on the tangent bundle of \mathbb{M} , $\mathcal{F} : T(\mathbb{M}) \rightarrow [0, +\infty]$, of which specific properties are discussed later, to define a geometry on \mathbb{M} . Any such Finsler metric \mathcal{F} induces a measure of length $\text{Length}_{\mathcal{F}}$ on the class of paths with Lipschitz regularity, defined as¹

$$\text{Length}_{\mathcal{F}}(\gamma) := \int_0^1 \mathcal{F}(\gamma(t), \dot{\gamma}(t)) dt,$$

¹In contrast to previous works [DGDHM16, BDRS14, BDMS15, MDS⁺17, DBRS13] we parameterize such that the time integration stays on $[0, 1]$, and $t > 0$ is *not* a priori reserved (unless explicitly stated otherwise) for arc length parametrization (which satisfies $\mathcal{F}_{\gamma(t)}(\dot{\gamma}(t)) = 1$).

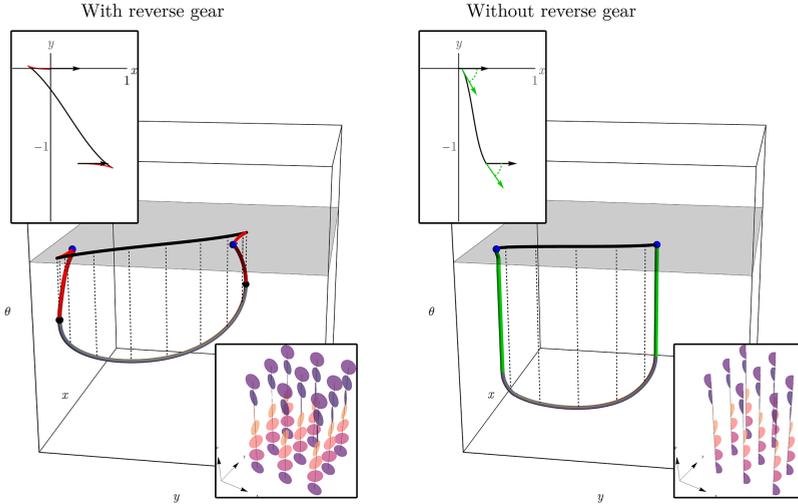


Figure 5.2: Top: Example of a shortest path with (left) and without (right) reverse gear in $\mathbb{R}^2 \times S$ and its projection on \mathbb{R}^2 . The black arrows indicate the begin and end condition in the plane, corresponding to the blue dots in $\mathbb{R}^2 \times S$. The paths in the lifted space are smooth, but vertical tangents appear in both cases. In the left figure, the projection of the path has two cusps, and the first and last part of the path is traversed backwards (the red parts). On the right, backward motion is not possible. Instead, according to our model, the shortest path is a concatenation of an in-place rotation (green), a SR geodesic, and again an in-place rotation. Bottom: corresponding control sets as defined in (5.7) for the allowed velocities at each position and orientation, with $B_{\mathcal{F}_0}$ on the left and $B_{\mathcal{F}_+}$ on the right.

with the convention $\dot{\gamma}(t) := \frac{d}{dt}\gamma(t)$. The path is said to be *normalized* w.r.t. \mathcal{F} iff $\mathcal{F}(\gamma(t), \dot{\gamma}(t)) = \text{Length}_{\mathcal{F}}(\gamma)$ for all $t \in [0, 1]$. Any Lipschitz continuous path of finite length can be normalized by a suitable reparametrization. Finally, the quasi-distance $d_{\mathcal{F}} : \mathbb{M} \times \mathbb{M} \rightarrow [0, +\infty]$ is defined for all $\mathbf{p}, \mathbf{q} \in \mathbb{M}$ by

$$d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) := \inf \{ \text{Length}_{\mathcal{F}}(\gamma) \mid \gamma \in \Gamma, \gamma(0) = \mathbf{p}, \gamma(1) = \mathbf{q} \}, \quad (5.1)$$

with $\Gamma := \text{Lip}([0, 1], \mathbb{M})$. Normalized minimizers of (5.1) are called *minimizing geodesics* from \mathbf{p} to \mathbf{q} w.r.t. \mathcal{F} . For certain pairs (\mathbf{p}, \mathbf{q}) these minimizers may not be unique, and these points are often of interest, see for example [MS10, BDMS17].

Definition 5.1.1 (Maxwell point). *Let $\mathbf{p}_S \in \mathbb{M}$ be a fixed point source and $\gamma \in \Gamma$ a geodesic connecting \mathbf{p}_S with $\mathbf{q} \in \mathbb{M}$, $\mathbf{q} \neq \mathbf{p}_S$. Then \mathbf{q} is a Maxwell point if there exists another extremal path $\tilde{\gamma} \in \Gamma$ connecting \mathbf{p}_S and \mathbf{q} , with $\text{Length}_{\mathcal{F}}(\gamma) = \text{Length}_{\mathcal{F}}(\tilde{\gamma})$. If \mathbf{q} is the first point (distinct from \mathbf{p}_S) on γ where such $\tilde{\gamma}$ exists, then \mathbf{q} is called the first Maxwell point. The curves $\gamma, \tilde{\gamma}$ lose global optimality after the first Maxwell point.*

Remark 5.1.2 (Terminology). *We use the common terminology of ‘Finsler metric’ for \mathcal{F} , although it is also called ‘Finsler function’, ‘Finsler norm’ or ‘Finsler structure’, and despite the fact that \mathcal{F} is not a metric (distance) in the classical sense. The Finsler metric \mathcal{F} induces the quasi-distance $d_{\mathcal{F}}$ as defined in (5.1). If $\mathcal{F}(\mathbf{p}, \dot{\mathbf{p}}) = \mathcal{F}(\mathbf{p}, -\dot{\mathbf{p}})$ for all $\mathbf{p} \in \mathbb{M}$ and tangent vectors $\dot{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{M})$, then $d_{\mathcal{F}}$ is a true metric, satisfying $d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) = d_{\mathcal{F}}(\mathbf{q}, \mathbf{p})$ for all $\mathbf{p}, \mathbf{q} \in \mathbb{M}$. However, to avoid confusion of the word metric, we will only refer to $d_{\mathcal{F}}$ as a distance or quasi-distance. Throughout the document, we use the words path and curve synonymously. When we consider the formal curve optimization problem (5.1), we speak of geodesics for the stationary curves. Such stationary curves are locally minimizing. A global minimizer of (5.1) is referred to as minimizing geodesic or minimizer.*

Remark 5.1.3. *Recall from Fig. 2.4 that the key advantage of bringing geodesic wave-front propagation to $\mathbb{R}^d \times S^{d-1}$ instead of \mathbb{R}^d is that the fronts can be prevented from ‘leaking’ at crossings of structures.*

5.1.2 Geometry of the Reeds-Shepp model

We introduce the Finsler metric \mathcal{F}_0 underlying the Reeds-Shepp car model, and the Finsler metric \mathcal{F}_0^+ corresponding to the variant without reverse gear. Let $(\mathbf{p}, \dot{\mathbf{p}}) \in T(\mathbb{M})$ be a pair consisting of a point $\mathbf{p} \in \mathbb{M}$ and a tangent vector $\dot{\mathbf{p}} \in T_{\mathbf{p}}(\mathbb{M})$ at this point. The physical and angular components of a point $\mathbf{p} \in \mathbb{M}$ are denoted by $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{n} \in S^{d-1}$, and this convention carries over to the tangent:

$$\mathbf{p} = (\mathbf{x}, \mathbf{n}) \in \mathbb{M}, \quad \dot{\mathbf{p}} = (\dot{\mathbf{x}}, \dot{\mathbf{n}}) \in T_{\mathbf{p}}(\mathbb{M}).$$

We say that $\dot{\mathbf{x}}$ is proportional to \mathbf{n} , that we write as $\dot{\mathbf{x}} \propto \mathbf{n}$, iff there exists a $\lambda \in \mathbb{R}$ such that $\dot{\mathbf{x}} = \lambda \mathbf{n}$. Define

$$\mathcal{F}_0(\mathbf{p}, \dot{\mathbf{p}})^2 := \begin{cases} \mathcal{C}_1^2(\mathbf{p})|\dot{\mathbf{x}} \cdot \mathbf{n}|^2 + \mathcal{C}_2^2(\mathbf{p})\|\dot{\mathbf{n}}\|^2 & \text{if } \dot{\mathbf{x}} \propto \mathbf{n}, \\ +\infty & \text{otherwise.} \end{cases} \quad (5.2)$$

$$\mathcal{F}_0^+(\mathbf{p}, \dot{\mathbf{p}})^2 := \begin{cases} \mathcal{C}_1^2(\mathbf{p})|\dot{\mathbf{x}} \cdot \mathbf{n}|^2 + \mathcal{C}_2^2(\mathbf{p})\|\dot{\mathbf{n}}\|^2 & \text{if } \dot{\mathbf{x}} \propto \mathbf{n} \text{ and } \dot{\mathbf{x}} \cdot \mathbf{n} \geq 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (5.3)$$

Here $\|\cdot\|$ denotes the norm and “ \cdot ” the usual inner product on the Euclidean space \mathbb{R}^d . The functions \mathcal{C}_1 and \mathcal{C}_2 are assumed to be continuous on \mathbb{M} , and uniformly bounded from below by a positive constant $\delta > 0$. In applications, \mathcal{C}_1 and \mathcal{C}_2 are chosen so as to favor paths which remain close to regions of interest, e.g. along blood vessels in retinal images, see Fig. 5.1. Note that their physical units are distinct: if one wishes $d_{\mathcal{F}}$ to

have the dimension $[T]$ of a travel time, then C_1^{-1} is a physical, (strictly) spatial velocity $[\text{Length}][T]^{-1}$, and C_2^{-1} is an angular velocity $[\text{Rad}][T]^{-1}$. For simplicity one often sets $C_1 = \xi C_2$, where $\xi^{-1} > 0$ is a unit of spatial length. The special case $C_1(\mathbf{p}) = \xi C_2(\mathbf{p}) = \xi$ for all $\mathbf{p} \in \mathbb{M}$ is referred to as the uniform cost case.

5.1.3 The eikonal equation and the fast marching algorithm

We compute the distance map to a point source on a volume using the relation to eikonal equations. Let $\mathbf{p}_S \in \mathbb{M}$ be an arbitrary source point, and let U be the associated distance function

$$U(\mathbf{p}) := d_{\mathcal{F}}(\mathbf{p}_S, \mathbf{p}). \quad (5.4)$$

Then U is the unique viscosity solution [CL83, CIL92] to the eikonal PDE:

$$\begin{cases} \mathcal{F}^*(\mathbf{p}, dU(\mathbf{p})) = 1 & \text{for all } \mathbf{p} \in \mathbb{M} \setminus \{\mathbf{p}_S\}, \\ U(\mathbf{p}_S) = 0. \end{cases} \quad (5.5)$$

Here \mathcal{F}^* is the dual metric of \mathcal{F} and dU is the differential of the distance map U . However, for these relations to hold, and for numerical discretization to be practical, \mathcal{F} should be at least continuous². We therefore propose in Section 5.2.3 for both \mathcal{F}_0 and \mathcal{F}_0^+ an approximating metric, that we denote by \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$, respectively, that are continuous and converge to \mathcal{F}_0 and \mathcal{F}_0^+ as $\varepsilon \rightarrow 0$. The approximating metrics correspond to a highly anisotropic Riemannian and Finslerian metric, rather than a sub-Riemannian or sub-Finslerian metric. The metric \mathcal{F}_ε is in line with previous approximations [CS06, BDMS15, SBD⁺15] for the case $d = 2$.

We design a monotone and causal discretization scheme for the static Hamilton-Jacobi PDE (5.5), which allows to apply an efficient, single pass Fast-Marching Algorithm [Tsi95]. Let us emphasize that designing a causal discretization scheme for (5.5) is non-trivial, because its local connectivity needs to obey an *acuteness property* [SV01, Vla06] depending on the geometry defined by \mathcal{F} . We provide constructions for the metrics \mathcal{F}_ε or $\mathcal{F}_\varepsilon^+$ of interest, based on the earlier works [Mir14, Mir13].

5.1.4 Contributions and outline

The extension to 3D of the Reeds-Shepp car model and the adaptation to model shortest paths for cars that cannot move backwards are new and provide an interesting collection of new theoretical and practical results:

²From a theoretical standpoint, one may rely on the notion of discontinuous viscosity solution [BCD97]. But this concept is outside of the scope of this chapter, and in addition it forbids the use of a singleton $\{\mathbf{p}_S\}$ as the target set.

- In Theorem 5.2.2 we show that the Reeds-Shepp model is globally and locally controllable, and that the Reeds-Shepp model without reverse gear is globally but not locally controllable. Hence the distance map loses continuity.
- We introduce regularizations \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$ of the Finsler metrics \mathcal{F}_0 and \mathcal{F}_0^+ , which make our numerical discretization possible. We show that both the corresponding distances converge to $d_{\mathcal{F}_0}$ and $d_{\mathcal{F}_0^+}$ as $\varepsilon \rightarrow 0$ and the minimizing curves converge to the ones for $\varepsilon = 0$, see Theorem 5.2.3. The fact that the limit distances are still viscosity solutions to the limit HJ equation coincides with the results of [Dra07].
- We present and prove for $d = 2$ and uniform cost a theorem that describes the occurrence of cusps for the sub-Riemannian model using \mathcal{F}_0 , and that using \mathcal{F}_0^+ leads to geodesics that are a concatenation of purely angular motion, a sub-Riemannian geodesic without cusps and again a purely angular motion. We call the positions where in-place rotation (or purely angular motion) takes place *keypoints*. For uniform cost, we show that the only possible keypoints are the begin and end point, and for many end conditions we can describe how this happens. The precise theoretical statement and proof are found in Theorem 5.2.10.
- Furthermore, we show in Theorem 5.2.15 how the geodesics can be obtained from the distance map, for a general Finsler metric, and in the more specific cases that we use in this chapter. For our cases of interest, we show that backtracking of geodesics is either done via a single intrinsic gradient descent (for the models with reverse gear), or via two intrinsic gradient descents (for the model without reverse gear).
- For our numerical experiments we make use of a Fast-Marching implementation, for $d = 2$ introduced in [Mir14]. In Section 6 we give a summary of the numerical approach for $d = 3$, but a detailed discussion of the implementation and an evaluation of the accuracy of the method is beyond the scope of this chapter. For $d = 2$, we show an extensive comparison between the models with and without reverse gear for uniform cost, to illustrate the useful principle of the keypoints, and to show the qualitative difference between the two models.

Outline In Section 5.2, we give a detailed overview of the theoretical results of the chapter. The theorems 5.2.2, 5.2.10 and 5.2.15 are discussed and proven in Sections 5.3, 5.4 and 5.5, respectively. The proof of Theorem 5.2.3 is given in Appendix B.1. We discuss the numerics briefly in Section 5.6. Conclusion and discussion follow in Section 5.7. The reader who is primarily interested in the application of these methods may choose to turn to Chapter 7.

5.2 Main results

In this section, we state formally the mathematical results announced in Section 5.1. Some preliminaries regarding the distance function are introduced in the Section below. Results regarding the exact Reeds-Shepp car models are gathered in Section 5.2.2. The description of the approximate models and the related convergence results appear in Section 5.2.3. Analysis of special interest points (cusps and keypoints) are done in Section 5.2.4. Results on the eikonal equation, and subsequent backtracking of minimizing geodesics via intrinsic gradients is presented in Section 5.2.5.

5.2.1 Preliminaries on the (quasi-)distance function and underlying geometry

Geometries on the manifold of states $\mathbb{M} = \mathbb{R}^d \times S^{d-1}$ are defined by means of Finsler metrics which are functions $\mathcal{F} : T(\mathbb{M}) \rightarrow [0, +\infty]$. On each tangent space, the metric should be 1-homogeneous, convex and quantitatively non-degenerate with a uniform constant $\delta > 0$: for all $\mathbf{p} = (\mathbf{x}, \mathbf{n}) \in \mathbb{M}$, $\dot{\mathbf{p}}, \dot{\mathbf{p}}_0, \dot{\mathbf{p}}_1 \in T_{\mathbf{p}}(\mathbb{M})$, and $\lambda \geq 0$:

$$\begin{aligned} \mathcal{F}(\mathbf{p}, \lambda \dot{\mathbf{p}}) &= \lambda \mathcal{F}(\mathbf{p}, \dot{\mathbf{p}}), & \mathcal{F}(\mathbf{p}, \dot{\mathbf{p}}_0 + \dot{\mathbf{p}}_1) &\leq \mathcal{F}(\mathbf{p}, \dot{\mathbf{p}}_0) + \mathcal{F}(\mathbf{p}, \dot{\mathbf{p}}_1), \\ \mathcal{F}(\mathbf{p}, \dot{\mathbf{p}}) &\geq \delta \sqrt{\|\dot{\mathbf{x}}\|^2 + \|\dot{\mathbf{n}}\|^2}. \end{aligned} \quad (5.6)$$

A weak regularity property is required as well, see the next remark. The induced distance $d_{\mathcal{F}}$, defined in (5.1), obeys $d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) = 0$ iff $\mathbf{p} = \mathbf{q}$, and obeys the triangle inequality. However, unlike a regular distance, $d_{\mathcal{F}}$ needs not be finite, or continuous, or symmetric in its arguments. Note that \mathcal{F}_0 and \mathcal{F}_0^+ as defined in (5.2) and (5.3), respectively, indeed satisfy the properties in (5.6).

Remark 5.2.1. *In contrast to the more common definition of Finsler metrics, we will not assume the Finsler metric to be smooth on $T(\mathbb{M})$, but use a weaker condition instead. Following [Che16], we require that the sets*

$$\mathcal{B}_{\mathcal{F}}(\mathbf{p}) := \{\dot{\mathbf{p}} \in T_{\mathbf{p}}\mathbb{M} \mid \mathcal{F}(\mathbf{p}, \dot{\mathbf{p}}) \leq 1\} \quad (5.7)$$

are closed and vary continuously with respect to the point $\mathbf{p} \in \mathbb{M}$ in the sense of the Hausdorff distance. The sets $\mathcal{B}_{\mathcal{F}}(\mathbf{p})$ are illustrated in Fig. 5.2 for the models of interest. The condition implies that a shortest path exists from \mathbf{p} to $\mathbf{q} \in \mathbb{M}$ whenever $d_{\mathcal{F}}(\mathbf{p}, \mathbf{q})$ is finite, and is used to prove convergence results in Appendix B.1.

A common technique in optimal control theory is to reformulate the shortest path problem defining the distance $d_{\mathcal{F}}(\mathbf{p}, \mathbf{q})$ into a time optimal control problem. That is,

for $p \in [1, \infty]$ one has by Hölder's (in)equality, time re-parametrization, and by 1-homogeneity of \mathcal{F} in its 2nd entry, that:

$$\begin{aligned} d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) &= \inf \left\{ \int_0^1 \mathcal{F}(\gamma(t), \dot{\gamma}(t)) dt \mid \gamma \in \Gamma, \gamma(0) = \mathbf{p}, \gamma(1) = \mathbf{q} \right\} \\ &= \inf \left\{ \left(\int_0^1 |\mathcal{F}(\gamma(t), \dot{\gamma}(t))|^p dt \right)^{\frac{1}{p}} \mid \gamma \in \Gamma, \gamma(0) = \mathbf{p}, \gamma(1) = \mathbf{q} \right\} \end{aligned} \quad (5.8)$$

$$= \inf \{ T \geq 0 \mid \exists \gamma \in \Gamma_T, \gamma(0) = \mathbf{p}, \gamma(T) = \mathbf{q}, \forall t \in [0, T] \dot{\gamma}(t) \in \mathcal{B}_{\mathcal{F}}(\gamma(t)) \}, \quad (5.9)$$

where $\Gamma_T := \text{Lip}([0, T], \mathbb{M})$, and with $\mathcal{B}_{\mathcal{F}}(\mathbf{p})$ as defined in (5.7). The latter reformulation is used in Appendix A to prove convergence results via closedness of controllable paths and Arzela-Ascoli's theorem, based on a general result originally applied to Euler elastica curves in [Che16].

In the special case $\mathcal{F} = \mathcal{F}_0$ the geodesics are SR geodesics, where \mathcal{F}_0 is obtained by the square root of quadratic form associated to a SR metric $\mathcal{G}_0|_{\mathbf{p}}(\cdot, \cdot) = \mathcal{F}_0(\mathbf{p}, \cdot)^2$ on a SR manifold $(\mathbb{M}, \Delta, \mathcal{G}_0)$, where $\Delta \subset T(\mathbb{M})$ is a strict subset of allowable tangent vectors that comes along with the horizontality constraint

$$\dot{\mathbf{x}}(t) = (\dot{\mathbf{x}}(t) \cdot \mathbf{n}(t))\mathbf{n}(t), \quad \forall t \in [0, 1], \quad (5.10)$$

that arises from (5.2). For details on the case $d = 2$ see [BDRS14, Sac11], for $d = 3$ see [DGDHM16].

Finally, we note that for the uniform cost case ($\xi^{-1}\mathcal{C}_1 = \mathcal{C}_2 = 1$), the problem is covariant with respect to rotations and translations. For the data-driven case, such covariance is only obtained when simultaneously rotating the data-driven cost factors $\mathcal{C}_1, \mathcal{C}_2$. Therefore, only in the uniform cost case, for $d = 2, 3$, we shall use a reference point ('the origin') $\mathbf{e} \in \mathbb{R}^d \times S^{d-1}$. To adhere to common conventions we use

$$\begin{aligned} \mathbf{e} = (\mathbf{0}, \mathbf{a}) \in \mathbb{R}^d \times S^{d-1}, \text{ with } \mathbf{a} := (1, 0)^T \text{ if } d = 2 \quad \text{and} \\ \mathbf{a} := (0, 0, 1)^T \text{ if } d = 3. \end{aligned} \quad (5.11)$$

5.2.2 Controllability of the Reeds-Shepp model

A model $(\mathbb{M}, d_{\mathcal{F}})$ is *globally controllable* if the distance $d_{\mathcal{F}}$ takes finite values on $\mathbb{M} \times \mathbb{M}$, in other words, a car can go from any place on the manifold to any other place in finite time. In Theorem 5.2.2 we show that this is indeed the case for $\mathcal{F} = \mathcal{F}_0$ and $\mathcal{F} = \mathcal{F}_0^+$, given in (5.2) and (5.3). *Local controllability* is satisfied when $d_{\mathcal{F}}$ satisfies a certain continuity requirement: if $\mathbf{p} \rightarrow \mathbf{q} \in (\mathbb{M}, \|\cdot\|)$, with $\|\cdot\|$ denoting the standard (flat) Euclidean norm on $\mathbb{M} = \mathbb{R}^d \times S^{d-1}$, we must have $d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) \rightarrow 0$. We prove in Theorem 5.2.2 that the metric space $(\mathbb{M}, d_{\mathcal{F}_0})$ is locally controllable, but the quasi-metric space $(\mathbb{M}, d_{\mathcal{F}_0^+})$ is not.

Indeed the SR Reeds-Shepp car can achieve sideways motions by alternating the forward and reverse gear with slight direction changes, whereas the model without reverse gear lacks this possibility. For completeness, the theorem contains a standard (rough) estimate of the distance near the source (due to well-known estimates [Gro96, tER98, CS06], see also Chapter 4.

Furthermore, we prove existence of minimizers for the Reeds-Shepp model without reverse gear. Existence results of minimizers of the model with reverse gear (the SR model) already exist, by the Chow-Rashevski theorem and Fillipov theorems [AS04].

Theorem 5.2.2 ((Local) controllability properties). *Minimizers exist for both the classical Reeds-Shepp model, and for the Reeds-Shepp model without reverse gear. Both models are globally controllable.*

- The Reeds-Shepp model without reverse gear is not locally controllable, since

$$\limsup_{\mathbf{p}' \rightarrow \mathbf{p}} d_{\mathcal{F}_0^+}(\mathbf{p}, \mathbf{p}') \geq 2\pi\delta, \text{ for all } \mathbf{p} \in \mathbb{M}, \quad (5.12)$$

with $\delta > 0$. If the cost $\mathcal{C}_2 = \delta$ is constant on \mathbb{M} , then this inequality is sharp:

$$\limsup_{\mathbf{p}' \rightarrow \mathbf{p}} d_{\mathcal{F}_0^+}(\mathbf{p}, \mathbf{p}') = \lim_{\mu \downarrow 0} d_{\mathcal{F}_0^+}((\mathbf{x}, \mathbf{n}), (\mathbf{x} - \mu\mathbf{n}, \mathbf{n})) = 2\pi\delta. \quad (5.13)$$

- The sub-Riemannian Reeds-Shepp model is locally controllable, since

$$d_{\mathcal{F}_0}(\mathbf{p}, \mathbf{p}') = \mathcal{O}\left(\mathcal{C}_2(\mathbf{p})\|\mathbf{n} - \mathbf{n}'\| + \sqrt{\mathcal{C}_2(\mathbf{p})\mathcal{C}_1(\mathbf{p})\|\mathbf{x} - \mathbf{x}'\|}\right) \\ \text{as } \mathbf{p}' = (\mathbf{x}', \mathbf{n}') \rightarrow \mathbf{p} = (\mathbf{x}, \mathbf{n}). \quad (5.14)$$

For a proof see Section 5.3.

5.2.3 A continuous approximation for the Reeds-Shepp geometry

We introduce approximations \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$ of the Finsler metrics \mathcal{F}_0 and \mathcal{F}_0^+ , depending on a small parameter $0 < \varepsilon \leq 1$, which are continuous and in particular take only finite values. This is a prerequisite for our numerical methods. Both approximations penalize the deviation from the constraints of collinearity $\dot{\mathbf{x}} \propto \mathbf{n}$, and in addition, $\mathcal{F}_\varepsilon^+$ penalizes negativity of the scalar product $\dot{\mathbf{x}} \cdot \mathbf{n}$, appearing in (5.2) and (5.3). For that purpose, we introduce some additional notation: for $\dot{\mathbf{x}} \in \mathbb{R}^d$ and $\mathbf{n} \in S^{d-1}$ we define

$$\|\dot{\mathbf{x}} \wedge \mathbf{n}\|^2 := \|\dot{\mathbf{x}}\|^2 - |\dot{\mathbf{x}} \cdot \mathbf{n}|^2, \quad (\dot{\mathbf{x}} \cdot \mathbf{n})_- := \min\{0, \dot{\mathbf{x}} \cdot \mathbf{n}\}, \quad (\dot{\mathbf{x}} \cdot \mathbf{n})_+ := \max\{\dot{\mathbf{x}} \cdot \mathbf{n}, 0\}. \quad (5.15)$$

These are respectively the norm of the orthogonal projection³ of $\dot{\mathbf{x}}$ onto the plane orthogonal to \mathbf{n} , and the negative and positive parts of their scalar product. The two metrics

³The quantity $\|\dot{\mathbf{x}} \wedge \mathbf{n}\|$ is also the norm of the wedge product of $\dot{\mathbf{x}}$ and \mathbf{n} , but defining it this way would require introducing some algebra which is not needed in the rest of this chapter.

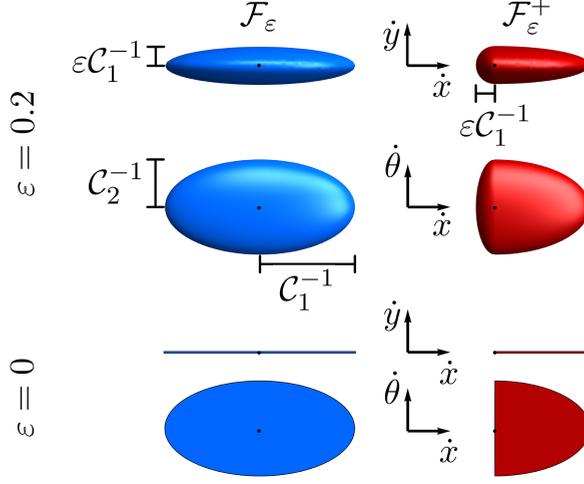


Figure 5.3: Levelsets for $d = 2$ of the (approximating) metrics $\mathcal{F}_\varepsilon(\mathbf{0}, (\dot{x}, \dot{y}, \dot{\theta})) = 1$ (left) and $\mathcal{F}_\varepsilon^+(\mathbf{0}, (\dot{x}, \dot{y}, \dot{\theta})) = 1$ (right), with $\varepsilon = 0.2$ (top) and $\varepsilon = 0$ (bottom). In this example, $\mathcal{C}_2(\mathbf{0}) = 2\mathcal{C}_1(\mathbf{0})$.

$\mathcal{F}_\varepsilon, \mathcal{F}_\varepsilon^+ : T(\mathbb{M}) \rightarrow \mathbb{R}^+$ are defined for each $0 < \varepsilon \leq 1$, as follows: for $(\mathbf{p}, \dot{\mathbf{p}}) \in T(\mathbb{M})$ with components $\mathbf{p} = (\mathbf{x}, \mathbf{n})$ and $\dot{\mathbf{p}} = (\dot{\mathbf{x}}, \dot{\mathbf{n}})$ we define

$$\mathcal{F}_\varepsilon(\mathbf{p}, \dot{\mathbf{p}})^2 := \mathcal{C}_1(\mathbf{p})^2 (|\dot{\mathbf{x}} \cdot \mathbf{n}|^2 + \varepsilon^{-2} \|\dot{\mathbf{x}} \wedge \mathbf{n}\|^2) + \mathcal{C}_2(\mathbf{p})^2 \|\dot{\mathbf{n}}\|^2, \quad (5.16)$$

$$\begin{aligned} \mathcal{F}_\varepsilon^+(\mathbf{p}, \dot{\mathbf{p}})^2 &:= \mathcal{C}_1(\mathbf{p})^2 (|\dot{\mathbf{x}} \cdot \mathbf{n}|^2 + \varepsilon^{-2} \|\dot{\mathbf{x}} \wedge \mathbf{n}\|^2 + (\varepsilon^{-2} - 1)(\dot{\mathbf{x}} \cdot \mathbf{n})_-^2) + \mathcal{C}_2(\mathbf{p})^2 \|\dot{\mathbf{n}}\|^2 \\ &= \mathcal{C}_1(\mathbf{p})^2 ((\dot{\mathbf{x}} \cdot \mathbf{n})_+^2 + \varepsilon^{-2} \|\dot{\mathbf{x}} \wedge \mathbf{n}\|^2 + \varepsilon^{-2} (\dot{\mathbf{x}} \cdot \mathbf{n})_-^2) + \mathcal{C}_2(\mathbf{p})^2 \|\dot{\mathbf{n}}\|^2. \end{aligned} \quad (5.17)$$

See Fig. 5.3 for a visualization of a level set of both metrics in $\mathbb{R}^2 \times S^1$. Note that \mathcal{F}_ε is a Riemannian metric on \mathbb{M} (with the same smoothness as the cost functions $\mathcal{C}_2, \mathcal{C}_1$), and that $\mathcal{F}_\varepsilon^+$ is neither Riemannian nor smooth due to the term $(\dot{\mathbf{x}} \cdot \mathbf{n})_-$. One clearly has the pointwise convergence $\mathcal{F}_\varepsilon(\mathbf{p}, \dot{\mathbf{p}}) \rightarrow \mathcal{F}_0(\mathbf{p}, \dot{\mathbf{p}})$ as $\varepsilon \rightarrow 0$, and likewise $\mathcal{F}_\varepsilon^+(\mathbf{p}, \dot{\mathbf{p}}) \rightarrow \mathcal{F}_0^+(\mathbf{p}, \dot{\mathbf{p}})$. The use of \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$ is further justified by the following convergence result.

Theorem 5.2.3 (Convergence of the Approximative Models to the Exact Models). *One has the pointwise convergence: for any $\mathbf{p}, \mathbf{q} \in \mathbb{M}$*

$$d_{\mathcal{F}_\varepsilon}(\mathbf{p}, \mathbf{q}) \rightarrow d_{\mathcal{F}_0}(\mathbf{p}, \mathbf{q}), \quad d_{\mathcal{F}_\varepsilon^+}(\mathbf{p}, \mathbf{q}) \rightarrow d_{\mathcal{F}_0^+}(\mathbf{p}, \mathbf{q}), \quad \text{as } \varepsilon \rightarrow 0.$$

Consider for each $\varepsilon > 0$ a minimizing path γ_ε^* from \mathbf{p} to \mathbf{q} , with respect to the metric \mathcal{F}_ε , parametrized at constant speed

$$\mathcal{F}_\varepsilon(\dot{\gamma}_\varepsilon^*(t), \dot{\gamma}_\varepsilon^*(t)) = d_{\mathcal{F}_\varepsilon}(\mathbf{p}, \mathbf{q}), \quad \forall t \in [0, 1].$$

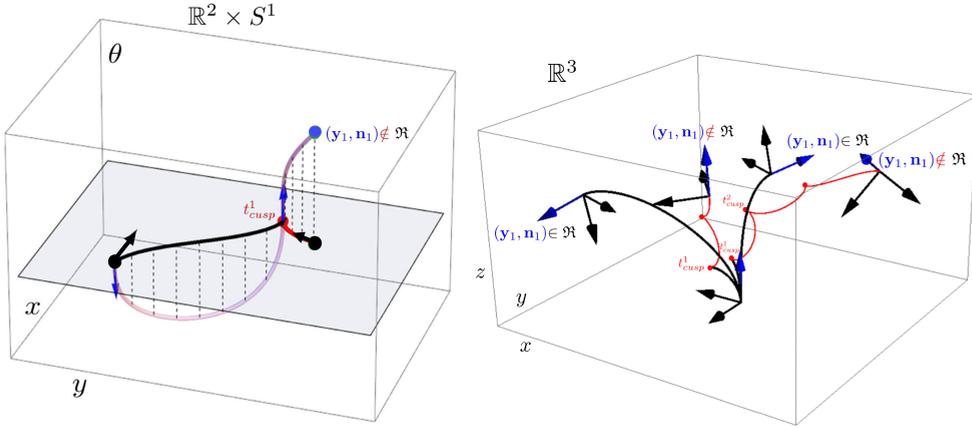


Figure 5.4: Illustration of cusps in SR ($\varepsilon = 0$) geodesics (possibly non-optimal) in $\mathbb{M} = \mathbb{R}^d \times S^{d-1}$. Left: cusps in spatial projections $\mathbf{x}(\cdot)$ of SR geodesics $\gamma(\cdot) = (\mathbf{x}(\cdot), \mathbf{n}(\cdot))$ for $d = 2$, right: cusps (red dots) appearing in spatial projections of SR geodesics for $d = 3$. In the 3D case we indicate the corresponding rotations $\mathbf{R}_{\mathbf{n}_1}$ via a local 3D frame.

Assume that there is a unique shortest path γ^* from \mathbf{p} to \mathbf{q} with respect to the sub-Riemannian distance $d_{\mathcal{F}_0}$ (in other words \mathbf{q} is not within the cut locus of \mathbf{p}), parametrized at constant speed:

$$\mathcal{F}_0(\gamma^*(t), \dot{\gamma}^*(t)) = d_{\mathcal{F}_0}(\mathbf{p}, \mathbf{q}), \quad \forall t \in [0, 1].$$

Then $\gamma_\varepsilon^* \rightarrow \gamma^*$ as $\varepsilon \rightarrow 0$, uniformly on $[0, 1]$. The same holds for $\mathcal{F}_\varepsilon^+$ for all $\varepsilon \geq 0$.

The proof, presented in Appendix B.1 is based on a general result originally applied to the Euler elastica curves in [Che16]. Combining Theorem 5.2.3 with the local controllability properties established in Theorem 5.2.2, one obtains that $d_{\mathcal{F}_\varepsilon} \rightarrow d_{\mathcal{F}_0}$ locally uniformly on $\mathbb{M} \times \mathbb{M}$, and that the convergence $d_{\mathcal{F}_\varepsilon^+} \rightarrow d_{\mathcal{F}_0^+}$ is only pointwise.

Remark 5.2.4. If there exists a family of minimizing geodesics $(\gamma_i^*)_{i \in I}$ from \mathbf{p} to \mathbf{q} with respect to \mathcal{F}_0 (resp. \mathcal{F}_0^+), then one can show that for any sequence $\varepsilon_n \rightarrow 0$ one can find a subsequence and an index $i \in I$ such that $\gamma_{\varepsilon_{\varphi(n)}}^* \rightarrow \gamma_i^*$ uniformly as $n \rightarrow \infty$.

5.2.4 Points of interest in spatial projections of geodesics for the uniform cost case: cusps vs. keypoints

Next we provide a theorem that tells us in each of the models/metric spaces $(\mathbb{M}, d_{\mathcal{F}_0})$, $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$ and $(\mathbb{M}, d_{\mathcal{F}_0^+})$, $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$, with $\mathcal{C}_1 = \mathcal{C}_2 = 1$ and $d = 2$ where cusps occur in spatial projections of geodesics or where keypoints with in-place rotations take place.

Recall Fig. 5.2 for a geometric illustration of the specific behavior of the path at such points. In Thm. 5.2.10, we provide an analysis of the occurrence of these points for the *uniform cost* case.

Note that for vessel-tracking (or fiber tracking) applications, cusps are not wanted as they are unnatural for vessels (or fibers), whereas keypoints are only desirable at bifurcations of vessels. In the data-driven case, the practical advantage of the forward-only model resulting in keypoints instead of cusps can indeed be observed (see e.g. Figs. 7.3 and 7.4).

Definition 5.2.5 (Cusp). *A cusp point $\mathbf{x}(t_0)$ on a spatial projection of a (SR) geodesic $t \mapsto (\mathbf{x}(t), \mathbf{n}(t))$ in \mathbb{M} is a point where*

$$\tilde{u}(t_0) = 0, \text{ and } \dot{\tilde{u}}(t_0) \neq 0, \text{ where } \tilde{u}(t) := \mathbf{n}(t) \cdot \dot{\mathbf{x}}(t) \text{ for all } t. \quad (5.18)$$

In other words, a cusp point is a point where the spatial control aligned with $\mathbf{n}(t_0)$ vanishes and switches sign locally.

Although this definition explains the notion of a cusp geometrically (as can be observed in Fig. 5.2 and Fig. 5.4), it contains a redundant part for the relevant case of interest: the second condition automatically follows when considering the SR geodesics in $(\mathbb{M}, d_{\mathcal{F}_0})$. The following lemma gives a characterization of a cusp point in terms of the distance function along a curve.

Lemma 5.2.6. *Consider a SR geodesic $\gamma = (\mathbf{x}, \mathbf{n}) : [0, 1] \rightarrow (\mathbb{M}, d_{\mathcal{F}_0})$, parametrized at constant speed, and which physical position $\mathbf{x}(\cdot)$ is not identically constant. Denote $\mathbf{p}_S := \gamma(0)$ and $U(\cdot) := d_{\mathcal{F}_0}(\mathbf{p}_S, \cdot)$. Let $t_0 \in (0, 1)$ be such that U is differentiable at $\gamma(t_0) = (\mathbf{x}(t_0), \mathbf{n}(t_0))$. Then*

$$\mathbf{x}(t_0) \text{ is a cusp point} \iff \mathbf{n}(t_0) \cdot \dot{\mathbf{x}}(t_0) = 0 \iff \mathbf{n}(t_0) \cdot \nabla_{\mathbb{R}^d} U(\mathbf{x}(t_0), \mathbf{n}(t_0)) = 0. \quad (5.19)$$

The proof can be found in Appendix B.3.

Definition 5.2.7 (Keypoint). *A point $\tilde{\mathbf{x}}$ on the spatial projection of a geodesic $\gamma(\cdot) = (\mathbf{x}(\cdot), \mathbf{n}(\cdot))$ in \mathbb{M} is a keypoint of γ if there exist $t_0 < t_1$, such that $\mathbf{x}(t) = \tilde{\mathbf{x}}$ and $\dot{\mathbf{n}}(t) \neq 0$ for all $t \in [t_0, t_1]$, i.e., a point where an in-place rotation takes place.*

Definition 5.2.8. *We define the set $\mathfrak{R} \subset \mathbb{M}$ to be all endpoints that can be reached with a geodesic $\gamma^* : [0, 1] \rightarrow \mathbb{M}$ in $(\mathbb{M}, d_{\mathcal{F}_0})$ whose spatial control $\tilde{u}(t)$ stays positive for all $t \in [0, 1]$.*

Remark 5.2.9. *The word ‘geodesic’ in this definition can (in the case $d = 2$) be replaced by ‘globally minimizing geodesic’ [BDRS14]. For a definition in terms of the exponential*

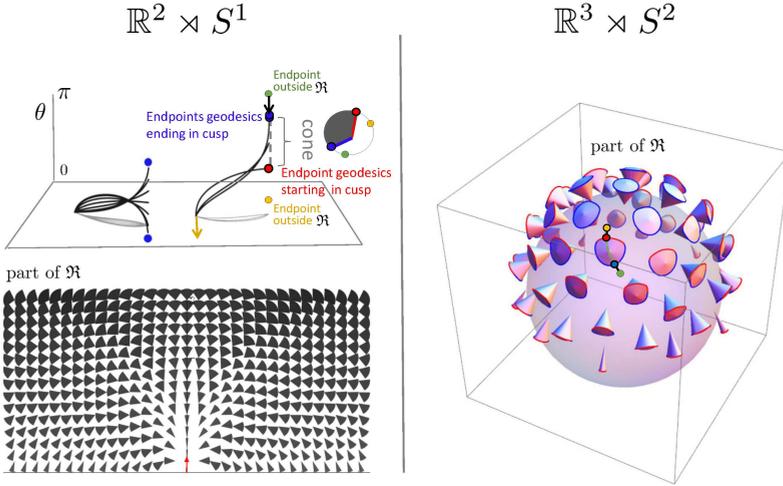


Figure 5.5: The set \mathfrak{R} of endpoints reachable from the origin \mathbf{e} (recall (5.11)) via SR geodesics whose spatial projections do not exhibit cusps has been studied for the case $d = 2$ (left), and for the case $d = 3$ (right). For $d = 2$ it is contained in $x \geq 0$ and for $d = 3$ it is contained in $z \geq 0$. The boundary of this set contains endpoints of geodesics departing at a cusp (in red) or of endpoints of geodesics ending in a cusp (in blue). If an endpoint (\mathbf{x}, \mathbf{n}) is placed outside \mathfrak{R} (e.g. the green points above) then following the approach in Theorem 5.2.15, depending on its initial spatial location it first connects to a blue point $(\mathbf{x}, \mathbf{n}_{new})$ via a spherical geodesic end then connects to the origin \mathbf{e} via a SR geodesic. Then it has a keypoint at the endpoint. For other spatial locations (orange points), the geodesic has the keypoint in the origin, or even at both boundaries, cf. Fig. 5.6.

map of a geometrical control problem \mathbf{P}_{curve} , see e.g. [DBRS13, DGDHS14], in which the same positivity condition for \tilde{u} is imposed. Fig. 5.5 shows more precisely what this set looks like for $d = 2$ [DBRS13] (in particular that it is contained in the half-space $\mathbf{a} \cdot \mathbf{x} \geq 0$, with \mathbf{a} as in (5.11)) and for $d = 3$ [DGDHS14]. We extend these results with the following theorem.

Theorem 5.2.10 (Cusps and Keypoints). *Let $\varepsilon > 0$, $d = 2$, $\mathcal{C}_1 = \mathcal{C}_2 = 1$. Then,*

- in $(\mathbb{M}, d_{\mathcal{F}_0})$ cusps are present in spatial projections of almost every optimal SR geodesics when their times t are extended to the real line (until they lose optimality). The straight lines connecting specific boundary points $\mathbf{p} = (\mathbf{x}, \mathbf{n})$ and $\mathbf{q} = (\mathbf{x} + \lambda \mathbf{n}, \mathbf{n})$ with $\lambda \in \mathbb{R}$ are the only exceptions.
- in $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$ and $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$ and $(\mathbb{M}, d_{\mathcal{F}_0^+})$ no cusps appear in spatial projections of geodesics.

Furthermore,

- in $(\mathbb{M}, d_{\mathcal{F}_0})$, $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$ and $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$ keypoints only occur with vertical geodesics (moving only angularly).
- in $(\mathbb{M}, d_{\mathcal{F}_0^+})$ keypoints only occur at the endpoints of shortest paths.

A minimizing geodesic γ_+ in $(\mathbb{M}, d_{\mathcal{F}_0^+})$ departing from $\mathbf{e} = (0, 0, 0)$ and ending in $\mathbf{p} = (x, y, \theta)$ has

- A) no keypoint if $\mathbf{p} \in \overline{\mathfrak{R}}$,
- B) a keypoint in $(0, 0)$ if $x < 0$,
- C) a keypoint only in (x, y) if⁴
 - C1) $\mathbf{p} \in \overline{\mathfrak{R}}^c$ and $x \geq 2$,
 - C2) $\mathbf{p} \in \overline{\mathfrak{R}}^c$ and $0 \leq x < 2$ and $|y| \leq -ix E\left(\frac{x}{\sqrt{4-x^2}}, \frac{x^2-4}{x^2}\right)$, where $E(z, m)$ denotes the Elliptic integral of the second kind.

We refer to Section 5.4 for the proof.

Remark 5.2.11. In case A, γ_+ is a minimizing geodesic in $(\mathbb{M}, d_{\mathcal{F}_0})$ as well. In case B, γ_+ departs from a cusp. In case C, γ_+ is a concatenation of a minimizing geodesic in $(\mathbb{M}, d_{\mathcal{F}_0})$ and an in-place rotation. For other endpoints (x, y, θ) for geodesics departing from \mathbf{e} with $0 \leq x < 2$, other than the ones reported in C2 it is not immediately clear what happens, due to [DBRS13, Thm.9]. Also points with $x < 0$ may have keypoints at the end as well. See Fig. 5.6 where various cases of minimizing geodesics in $(\mathbb{M}, d_{\mathcal{F}_0^+})$ are depicted.

Remark 5.2.12. See [DMMP16, Fig. 6] to see the smoothing effect of taking ε small but nonzero on the cusps of non-optimal geodesics in $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$ and keypoints in $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$.

5.2.5 The Eikonal PDE formalism

As briefly discussed in Sections 2.3 and 5.1.3, continuous metrics like \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$ for any $\varepsilon > 0$, allow to use the standard theory of viscosity solutions of eikonal PDEs, and thus to design provable and efficient numerical schemes for the computation of distance maps and minimizing geodesics. More precisely, consider a continuous Finsler metric $\mathcal{F} \in C^0(T(\mathbb{M}), \mathbb{R}^+)$, and define the dual \mathcal{F}^* on the co-tangent bundle as follows: for all $(\mathbf{p}, \hat{\mathbf{p}}) \in T^*(\mathbb{M})$

$$\mathcal{F}^*(\mathbf{p}, \hat{\mathbf{p}}) := \sup_{\dot{\mathbf{p}} \in T_{\mathbf{p}}\mathbb{M} \setminus \{0\}} \frac{\langle \hat{\mathbf{p}}, \dot{\mathbf{p}} \rangle}{\mathcal{F}(\mathbf{p}, \dot{\mathbf{p}})}. \tag{5.20}$$

⁴Here $\overline{\mathfrak{R}}^c = \mathbb{M} \setminus \overline{\mathfrak{R}}$ denotes the complement of the closure $\overline{\mathfrak{R}}$ of \mathfrak{R} , and $E(z, m) = \int_0^z \sqrt{1 - m \sin^2 v} dv$.

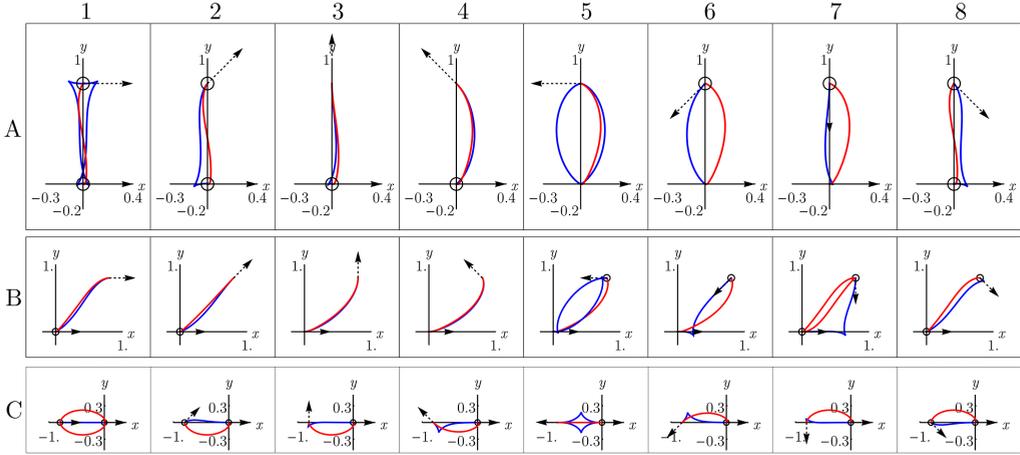


Figure 5.6: Shortest paths for $d = 2$ using the Finsler metrics \mathcal{F}_0 (blue) and \mathcal{F}_0^+ (red), with point source $\mathbf{p}_S = (0, 0, 0)$ and varying end conditions. Row A: $\mathbf{p} = (0, 0.8, \pi/4)$. Row B: $\mathbf{p} = (0.8, 0.8, \pi/4)$. Row C: $\mathbf{p} = (-0.8, 0, \pi/4)$. Here $n = 1, \dots, 8$, corresponding to the columns. When there are two minimizing geodesics, both are drawn. Circles around the begin or end point indicate in-place rotation of the red curve at that point. We see that whenever the blue geodesic has a cusp, the red geodesic has at least one in-place rotation (keypoint). This numerically supports our statements in Theorem 5.2.10 considering cusps and keypoints. For high accuracy we applied the relatively slow iterative PDE approach [BDMS15] on a $101 \times 101 \times 64$ -grid in \mathbb{M} to compute $d_{\mathcal{F}_0}(\mathbf{p}, \mathbf{p}_S)$ and $d_{\mathcal{F}_0^+}(\mathbf{p}, \mathbf{p}_S)$, see [DMMP16, App. B].

The distance map $U = d_{\mathcal{F}}(\mathbf{p}_S, \cdot)$ from a given source point $\mathbf{p}_S \in \mathbb{M}$ is the unique solution, in the sense of viscosity solutions, of the static Hamilton Jacobi equation: $U(\mathbf{p}_S) = 0$, and for all $p \in \mathbb{M}$

$$\mathcal{F}^*(\mathbf{p}, dU(\mathbf{p})) = 1. \tag{5.21}$$

Furthermore, if γ is a minimizing geodesic from \mathbf{p}_S to some $\mathbf{p} \in \mathbb{M}$, then it obeys the ordinary differential equation (ODE):

$$\begin{cases} \dot{\gamma}(t) = L \, d_{\hat{\mathbf{p}}} \mathcal{F}^*(\gamma(t), dU(\gamma(t))), \quad L := d_{\mathcal{F}}(\mathbf{p}_S, \mathbf{p}) \\ \gamma(0) = \mathbf{p}_S, \quad \gamma(1) = \mathbf{p}. \end{cases} \tag{5.22}$$

for any $t \in [0, 1]$ such that the differentiability of U and \mathcal{F}^* holds at the required points. The proof of the ODE (5.22) is for completeness derived in Proposition B.2.2 of Appendix B.2, where we also discuss in Remark B.2.3 the common alternative formalism based on the Hamiltonian. We denoted by $d_{\hat{\mathbf{p}}} \mathcal{F}^*$ the differential of the dual Finsler metric \mathcal{F}^* with respect to the second variable $\hat{\mathbf{p}}$, hence $d_{\hat{\mathbf{p}}} \mathcal{F}^*(\mathbf{p}, \hat{\mathbf{p}}) \in T_{\mathbf{p}}^{**}(\mathbb{M}) \cong T_{\mathbf{p}}(\mathbb{M})$ is indeed a tangent vector to \mathbb{M} , for all $(\mathbf{p}, \hat{\mathbf{p}}) \in T^*\mathbb{M}$.

In the rest of this section, we specialize (5.21) and (5.22) to the Finsler metrics \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$. Our first result provides explicit expressions for the dual Finsler metrics (required for the eikonal equation).

Proposition 5.2.13. *For any $0 < \varepsilon \leq 1$, the duals to the approximating Finsler metrics \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$ are: for all $(\mathbf{p}, \hat{\mathbf{p}}) \in T^*(\mathbb{M})$, with $\mathbf{p} = (\mathbf{x}, \mathbf{n})$ and $\hat{\mathbf{p}} = (\hat{\mathbf{x}}, \hat{\mathbf{n}})$*

$$\begin{aligned} \mathcal{F}_\varepsilon^*(\mathbf{p}, \hat{\mathbf{p}})^2 &= (\mathcal{C}_2(\mathbf{p}))^{-2} \|\hat{\mathbf{n}}\|^2 + (\mathcal{C}_1(\mathbf{p}))^{-2} (|\hat{\mathbf{x}} \cdot \mathbf{n}|^2 + \varepsilon^2 \|\hat{\mathbf{x}} \wedge \mathbf{n}\|^2) \\ \mathcal{F}_\varepsilon^{+*}(\mathbf{p}, \hat{\mathbf{p}})^2 &= (\mathcal{C}_2(\mathbf{p}))^{-2} \|\hat{\mathbf{n}}\|^2 + (\mathcal{C}_1(\mathbf{p}))^{-2} (|\hat{\mathbf{x}} \cdot \mathbf{n}|^2 + \varepsilon^2 \|\hat{\mathbf{x}} \wedge \mathbf{n}\|^2 - (1 - \varepsilon^2)(\hat{\mathbf{x}} \cdot \mathbf{n})_-^2) \\ &= (\mathcal{C}_2(\mathbf{p}))^{-2} \|\hat{\mathbf{n}}\|^2 + (\mathcal{C}_1(\mathbf{p}))^{-2} ((\hat{\mathbf{x}} \cdot \mathbf{n})_+^2 + \varepsilon^2 (\hat{\mathbf{x}} \cdot \mathbf{n})_-^2 + \varepsilon^2 \|\hat{\mathbf{x}} \wedge \mathbf{n}\|^2) \end{aligned} \quad (5.23)$$

In order to relate the Finslerian HJB equation (5.21) and backtracking equation (5.22) to some more classical Riemannian counterparts, we introduce two Riemannian metric tensor fields on \mathbb{M} . The first is defined as the polarization of the norm $\mathcal{F}_\varepsilon(\mathbf{p}, \cdot)$

$$\mathcal{G}_{\mathbf{p};\varepsilon}(\dot{\mathbf{p}}, \dot{\mathbf{p}}) = |\mathcal{F}_\varepsilon(\mathbf{p}, \dot{\mathbf{p}})|^2 = \mathcal{C}_1^2(\mathbf{p})((\dot{\mathbf{x}} \cdot \mathbf{n})^2 + \varepsilon^{-2} \|\dot{\mathbf{x}} \wedge \mathbf{n}\|^2) + \mathcal{C}_2^2(\mathbf{p}) \|\dot{\mathbf{n}}\|^2, \quad (5.24)$$

where $\dot{\mathbf{p}} = (\dot{\mathbf{x}}, \dot{\mathbf{n}})$, and then one can also rely on gradient fields $\mathbf{p} \mapsto \mathcal{G}_{\mathbf{p};\varepsilon}^{-1} dU(\mathbf{p})$ relative to this metric tensor. This has benefits if it comes to geometric understanding of the eikonal equation and its tracking. Even in the analysis of the non-symmetric case –where one does not have a single metric tensor– this notion plays a role, as we will see in the next main theorem. To this end, in the non-symmetric case, we shall rely on a second spatially isotropic metric tensor given by:

$$\tilde{\mathcal{G}}_{\mathbf{p};\varepsilon}(\dot{\mathbf{p}}, \dot{\mathbf{p}}) := \mathcal{C}_1^2(\mathbf{p}) \varepsilon^{-2} \|\dot{\mathbf{x}}\|^2 + \mathcal{C}_2^2(\mathbf{p}) \|\dot{\mathbf{n}}\|^2. \quad (5.25)$$

We denote by $\nabla_{S^{d-1}}$ the gradient operator on S^{d-1} with respect to the inner product induced by the embedding $S^{d-1} \subset \mathbb{R}^d$, and by $\nabla_{\mathbb{R}^d}$ the canonical gradient on \mathbb{R}^d .

Corollary 5.2.14. *Let $\varepsilon \geq 0$. Then the eikonal PDE (5.5) for the case $(\mathbb{M}, \mathcal{F}_\varepsilon)$ takes the form*

$$\begin{aligned} \sqrt{\frac{\|\nabla_{S^{d-1}} U(\mathbf{p})\|^2}{\mathcal{C}_2^2(\mathbf{p})} + \frac{\varepsilon^2 \|\nabla_{\mathbb{R}^d} U(\mathbf{p})\|^2 + (1 - \varepsilon^2) |\mathbf{n} \cdot \nabla_{\mathbb{R}^d} U(\mathbf{p})|^2}{\mathcal{C}_1^2(\mathbf{p})}} = 1 &\iff \\ \mathcal{G}_{\mathbf{p};\varepsilon}|_{\mathbf{p}} \left(\mathcal{G}_{\mathbf{p};\varepsilon}^{-1} dU(\mathbf{p}), \mathcal{G}_{\mathbf{p};\varepsilon}^{-1} dU(\mathbf{p}) \right) = 1. \end{aligned}$$

The eikonal PDE (5.5) for the case $(\mathbb{M}, \mathcal{F}_\varepsilon^+)$ now takes the explicit form:

$$\begin{aligned} \sqrt{\frac{\|\nabla_{S^{d-1}} U^+(\mathbf{p})\|^2}{\mathcal{C}_2^2(\mathbf{p})} + \frac{\varepsilon^2 \|\nabla_{\mathbb{R}^d} U^+(\mathbf{p})\|^2 + (1 - \varepsilon^2) |(\mathbf{n} \cdot \nabla_{\mathbb{R}^d} U^+(\mathbf{p}))_+|^2}{\mathcal{C}_1^2(\mathbf{p})}} = 1 &\iff \\ \begin{cases} \mathcal{G}_{\mathbf{p};\varepsilon}|_{\mathbf{p}} \left(\mathcal{G}_{\mathbf{p};\varepsilon}^{-1} dU^+(\mathbf{p}), \mathcal{G}_{\mathbf{p};\varepsilon}^{-1} dU^+(\mathbf{p}) \right) = 1, & \text{if } \mathbf{p} \in \mathbb{M}_+ := \{\mathbf{p} \in \mathbb{M} \mid \langle dU^+(\mathbf{p}), \mathbf{n} \rangle > 0\}, \\ \tilde{\mathcal{G}}_{\mathbf{p};\varepsilon}|_{\mathbf{p}} \left(\tilde{\mathcal{G}}_{\mathbf{p};\varepsilon}^{-1} dU^+(\mathbf{p}), \tilde{\mathcal{G}}_{\mathbf{p};\varepsilon}^{-1} dU^+(\mathbf{p}) \right) = 1, & \text{if } \mathbf{p} \in \mathbb{M}_- := \{\mathbf{p} \in \mathbb{M} \mid \langle dU^+(\mathbf{p}), \mathbf{n} \rangle < 0\}. \end{cases} \end{aligned}$$

for those $\mathbf{p} \in \mathbb{M}_+ \cup \mathbb{M}_-$ where U^+ is differentiable.

The proof of Proposition 5.2.13 and Corollary 5.2.14 can be found in Section 5.5.

We finally specialize the geodesic ODE (5.22) to the models of interest. Note that for the model $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$, the backtracking switches between qualitatively distinct modes, respectively almost sub-Riemannian and almost purely angular, in the spirit of Thm. 5.2.10. For $\varepsilon > 0$ and $\mathbf{n} \in S^{d-1}$ let $\mathbf{D}_\mathbf{n}^\varepsilon$ denote the $d \times d$ symmetric positive definite matrix with eigenvalue 1 in the direction \mathbf{n} , and eigenvalue ε^2 in the orthogonal directions:

$$\mathbf{D}_\mathbf{n}^\varepsilon := \mathbf{n} \otimes \mathbf{n} + \varepsilon^2(\text{Id} - \mathbf{n} \otimes \mathbf{n}). \quad (5.26)$$

Theorem 5.2.15 (Backtracking). *Let $0 < \varepsilon < 1$. Let $\mathbf{p}_S \in \mathbb{M}$ be a source point. Let $U(\mathbf{p}) := d_{\mathcal{F}_\varepsilon}(\mathbf{p}, \mathbf{p}_S)$, $U^+(\mathbf{p}) := d_{\mathcal{F}_\varepsilon^+}(\mathbf{p}, \mathbf{p}_S)$ be distance maps from \mathbf{p}_S , w.r.t. the Finsler metric \mathcal{F}_ε , and $\mathcal{F}_\varepsilon^+$. Let $\gamma, \gamma^+ : [0, 1] \rightarrow \mathbb{M}$ be normalized geodesics of length L starting at \mathbf{p}_S in $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$ resp. $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$. Let time $t \in [0, 1]$.*

For the Riemannian approximation paths of the Reeds-Shepp car we have, provided that U is differentiable at $\gamma(t) = (\mathbf{x}(t), \mathbf{n}(t))$, that

$$\dot{\gamma}(t) = L \mathcal{G}_{\gamma(t); \varepsilon}^{-1} dU(\gamma(t)) \iff \begin{cases} \dot{\mathbf{n}}(t) &= L \mathcal{C}_2(\gamma(t))^{-1} \nabla_{S^{d-1}} U(\gamma(t)), \\ \dot{\mathbf{x}}(t) &= L \mathcal{C}_1(\gamma(t))^{-1} \mathbf{D}_{\mathbf{n}(t)}^\varepsilon \nabla_{\mathbb{R}^d} U(\gamma(t)). \end{cases} \quad (5.27)$$

For the approximation paths of the car without reverse gear we have, provided that U^+ is differentiable at $\gamma^+(t) = (\mathbf{x}^+(t), \mathbf{n}^+(t))$, that

$$\dot{\gamma}^+(t) = L \begin{cases} \mathcal{G}_{\gamma^+(t); \varepsilon}^{-1} dU^+(\gamma^+(t)) & \text{if } \gamma^+(t) \in \mathbb{M}_+, \\ \tilde{\mathcal{G}}_{\gamma^+(t); \varepsilon}^{-1} dU^+(\gamma^+(t)) & \text{if } \gamma^+(t) \in \mathbb{M}_-, \end{cases} \quad (5.28)$$

with $\tilde{\mathcal{G}}_{\mathbf{p}; \varepsilon}(\dot{\mathbf{p}}, \dot{\mathbf{p}})$ given by (5.25), with disjoint Riemannian manifold splitting $\mathbb{M} = \mathbb{M}_+ \cup \mathbb{M}_- \cup \partial\mathbb{M}_\pm$. Manifold \mathbb{M}_+ is equipped with metric tensor \mathcal{G}_ε , \mathbb{M}_- is equipped with metric tensor $\tilde{\mathcal{G}}_\varepsilon$ and

$$\partial\mathbb{M}_\pm := \overline{\mathbb{M}_+} \setminus \mathbb{M}_+ = \overline{\mathbb{M}_-} \setminus \mathbb{M}_- \quad (5.29)$$

denotes the transition surface (surface of keypoints).

Remark 5.2.16. *The general abstract formula (5.28) reflects that the backtracking in $(\mathbb{M}, \mathcal{F}^+)$ is a combined gradient descent flow on the distance map U^+ on a splitting of \mathbb{M} into two (symmetric) Riemannian manifolds. Its explicit form (likewise (5.27)) is*

$$\begin{cases} \dot{\mathbf{n}}^+(t) &= L \mathcal{C}_2(\gamma^+(t))^{-1} \nabla_{S^{d-1}} U^+(\gamma^+(t)), \\ \dot{\mathbf{x}}^+(t) &= L \begin{cases} \mathcal{C}_1(\gamma^+(t))^{-1} \mathbf{D}_{\mathbf{n}(t)}^\varepsilon \nabla_{\mathbb{R}^d} U^+(\gamma^+(t)) & \text{if } \gamma^+(t) \in \mathbb{M}_+, \\ \varepsilon^2 \mathcal{C}_1(\gamma^+(t))^{-1} \nabla_{\mathbb{R}^d} U^+(\gamma^+(t)) & \text{if } \gamma^+(t) \in \mathbb{M}_-, \end{cases} \end{cases} \quad (5.30)$$

Note that for the isotropic case $\varepsilon = 1$, \mathcal{F}_1 and \mathcal{F}_1^+ coincide and geodesics consist of straight lines $\mathbf{x}(\cdot)$ in \mathbb{R}^d and great circles $\mathbf{n}(\cdot)$ in S^d that do not influence each other.

Remark 5.2.17. In Theorem 5.2.15, we assumed distance maps U and U^+ to be differentiable along the path, which is not always the case. In points where the distance map is not differentiable, one can take any sub-gradient in the sub-differential $\partial U(\mathbf{p})$ in order to identify Maxwell points (and Maxwell strata). In particular, in SR geometry, the set of points where the squared distance function $(d_{\mathcal{F}_0}(\cdot, e))^2$ is smooth is open and dense in any compact subset of \mathbb{M} , see [ABB16, Thm. 11.15]. The points where it is non-smooth are rare and meaningful: they are either first Maxwell points, conjugate points or abnormal points. The last type does not appear here, because we have a 2-bracket generating distribution, see e.g. [DGDHM16, Remark 4] and [ABB16, Ch. 20.5.1.]. At points in the closure of the first Maxwell set, two geodesically equidistant wavefronts collide for the first time, see for example [BDMS15, Fig.3, Thm 3.2] for the case $d = 2$ and $\mathcal{C} = \mathcal{C}_1 = \mathcal{C}_2 = 1$. See also Fig. 5.6, where for some end conditions 2 optimally back-tracked geodesics end with the same length in such a first Maxwell point. The conjugate points are points where local optimality is lost, for a precise definition see e.g. [ABB16, Def. 8.43].

Remark 5.2.18. Recall the convergence result from Theorem 5.2.3, and the non-local-controllability for the model $(\mathbb{M}, d_{\mathcal{F}_0^+})$. From this we see that the convergence holds point-wise but not uniformly (otherwise the limit distance $d_{\mathcal{F}_0^+}$ was continuous). Nevertheless the shortest paths converge strongly as $\varepsilon \downarrow 0$, and we see that the spatial velocity tends to 0 in (5.30) if $\varepsilon \downarrow 0$ if $\gamma_\varepsilon^*(t) \in \mathbb{M}_-$. In the SR case $\varepsilon = 0$, the gradient flows themselves fit continuously and the interface $\partial \mathbb{M}_\pm$ is reached with $\dot{\mathbf{x}} \cdot \mathbf{n} = 0$ (and $\dot{\mathbf{x}} = 0$).

Theorem 5.2.15 can be extended to the SR case:

Corollary 5.2.19 (SR Backtracking). *Let the cost $\mathcal{C}_1, \mathcal{C}_2$ be smooth, let the source $\mathbf{p}_S \in \mathbb{M}$ and $\mathbf{p} \neq \mathbf{p}_S \in \mathbb{M}$ be such that they can be connected by a unique smooth minimizer γ_ε^* in $(\mathbb{M}, \mathcal{F}_\varepsilon)$ and γ_0^* in $(\mathbb{M}, \mathcal{F}_0)$, such that $\gamma_\varepsilon^*(t)$ is not a conjugate point for all $t \in [0, 1]$ and all sufficiently small $\varepsilon > 0$, say $\varepsilon < \varepsilon_0$, for some $\varepsilon_0 > 0$. Then defining $U_0 : \mathbf{q} \in \mathbb{M} \mapsto d_{\mathcal{F}_\varepsilon}(\mathbf{p}_S, \mathbf{q})$ one has*

$$\dot{\gamma}_0^*(t) = U_0(\mathbf{p}) \mathcal{G}_{\gamma_0^*(t); 0}^{-1} dU_0(\gamma_0^*(t)), \quad t \in [0, 1],$$

assuming U_0 is differentiable at $\gamma_0^*(t)$. In addition U_0 satisfies the SR eikonal equation:

$$\sqrt{\mathcal{G}_{\mathbf{p}; 0} (\mathcal{G}_{\mathbf{p}; 0}^{-1} dU_0(\mathbf{p}), \mathcal{G}_{\mathbf{p}; 0}^{-1} dU_0(\mathbf{p}))} = 1.$$

Proof. From our assumptions on \mathbf{p} and $\gamma_\varepsilon^*(t)$ for $\varepsilon < \varepsilon_0$, we have, recall Remark 5.2.17, that $(U_\varepsilon(\cdot))^2$ is differentiable at $\gamma_\varepsilon^*(t)$ for all $0 \leq t \leq 1$ and $0 \leq \varepsilon < \varepsilon_0$. This implies that U_ε is differentiable at $\{\gamma_\varepsilon^*(t) \mid 0 < t \leq 1\}$, for all $0 < \varepsilon < \varepsilon_0$.

From Theorem 5.2.3 we have pointwise convergence $U_\varepsilon(\mathbf{p}) \rightarrow U_0(\mathbf{p})$ and uniform convergence $\gamma_\varepsilon^* \rightarrow \gamma_0^*$ as $\varepsilon \downarrow 0$. Moreover, as γ_ε^* and γ_0^* are solutions of the canonical ODEs of Pontryagin's Maximum Principle, the trajectories are continuously depending on $\varepsilon > 0$, and so are the derivatives $\dot{\gamma}_\varepsilon^*$. As a result, we can apply the backtracking Theorem 5.2.15 for $\varepsilon > 0$ and take the limits:

$$\begin{aligned}
\dot{\gamma}_0^*(t) &= \lim_{\varepsilon \downarrow 0} \dot{\gamma}_\varepsilon^*(t) \\
&\stackrel{\text{Thm. 4}}{=} \lim_{\varepsilon \downarrow 0} U_\varepsilon(\mathbf{p}) (\mathcal{G}_{\gamma_\varepsilon^*(t); \varepsilon}^{-1} dU_\varepsilon)(\gamma_\varepsilon^*(t)) \\
&= U_0(\mathbf{p}) \left(\lim_{\varepsilon \downarrow 0} \mathcal{G}_{\gamma_\varepsilon^*(t); \varepsilon}^{-1} \right) \left(\lim_{\varepsilon \downarrow 0} (dU_\varepsilon(\gamma_\varepsilon^*(t))) \right) \\
&\stackrel{\text{Thm. 2}}{=} U_0(\mathbf{p}) \mathcal{G}_{\gamma_0^*(t); 0}^{-1} (dU_0)(\gamma_0^*(t)).
\end{aligned} \tag{5.31}$$

Furthermore,

$$1 = \lim_{\varepsilon \downarrow 0} \sqrt{\mathcal{G}_{\mathbf{p}; \varepsilon} (\mathcal{G}_{\mathbf{p}; \varepsilon}^{-1} dU_\varepsilon(\mathbf{p}), \mathcal{G}_{\mathbf{p}; \varepsilon}^{-1} dU_\varepsilon(\mathbf{p}))} = \sqrt{\mathcal{G}_{\mathbf{p}; 0} (\mathcal{G}_{\mathbf{p}; 0}^{-1} dU_0(\mathbf{p}), \mathcal{G}_{\mathbf{p}; 0}^{-1} dU_0(\mathbf{p}))}$$

where we recall Corollary 5.2.14. Here due to our assumptions, U_ε and U_0 are both differentiable at \mathbf{p} . Note that the limit for the inverse metric $\mathcal{G}_{\mathbf{p}; \varepsilon}^{-1}$ as $\varepsilon \downarrow 0$ exists, recall Cor. 5.2.14. \square

Now that we stated our 4 main theoretical results we will prove them in the subsequent sections (and Appendix B.1).

5.3 Controllability properties: proof of Theorem 5.2.2, and Maxwell-points in $(\mathbb{M}, d_{\mathcal{F}_0^+})$

(Global controllability) The two considered Reeds-Shepp models $(\mathbb{M}, d_{\mathcal{F}_0})$ and $(\mathbb{M}, d_{\mathcal{F}_0^+})$ are globally controllable, in the sense that the distances $d_{\mathcal{F}_0}$ and $d_{\mathcal{F}_0^+}$ take finite values on $\mathbb{M} \times \mathbb{M}$. This easily follows from the observation that any path $\mathbf{x} : [0, 1] \rightarrow \mathbb{R}^d$, which time derivative $\dot{\mathbf{x}} := \frac{d\mathbf{x}}{dt}$ is Lipschitz and non-vanishing, can be lifted into a path $\gamma : [0, 1] \rightarrow \mathbb{M}$ of finite length w.r.t. \mathcal{F}_0 and \mathcal{F}_0^+ , defined by $\gamma(t) := (\mathbf{x}(t), \dot{\mathbf{x}}(t)/\|\dot{\mathbf{x}}(t)\|)$ for all $t \in [0, 1]$. The fact that the infimum in (5.1) is actually a minimum for $\mathcal{F} = \mathcal{F}_0^+$ follows by Corollary B.1.8 in App. B.1 and (5.9), and the fact that the quasi-distances take finite values.

(Local controllability) In order to show that the model $(\mathbb{M}, d_{\mathcal{F}_0^+})$ is *not* locally controllable, we need the following lemma.

Lemma 5.3.1. *Let $\mathbf{n} : [0, \pi] \rightarrow \mathbb{S}^{d-1}$ be strictly 1-Lipschitz. Then $\int_0^\pi \mathbf{n}(0) \cdot \mathbf{n}(t) dt > 0$. Let $\mathbf{n} : \mathbb{R} \rightarrow \mathbb{S}^{d-1}$ be strictly 1-Lipschitz and 2π -periodic. Then all points $\mathbf{n}(t)$ lay in a common strict hemisphere. In particular $\mathbf{0} \notin \text{Hull}\{\mathbf{n}(t) \mid t \in [0, 2\pi]\}$.*

Proof. The Lipschitzness assumption implies $\mathbf{n}(0) \cdot \mathbf{n}(t) > \cos(t)$ for all $t \in (0, \pi]$ so $\int_0^\pi \mathbf{n}(0) \cdot \mathbf{n}(t) dt > 0$.

Let $\mathbf{n} : \mathbb{R} \rightarrow \mathbb{S}^{d-1}$ be strictly 1-Lipschitz and 2π -periodic. Set $\mathbf{M} := \int_0^{2\pi} \mathbf{n}(t) dt$. Then for any $t_0 \in [0, 2\pi]$ one has by the two assumptions

$$\mathbf{n}(t_0) \cdot \mathbf{M} = \int_0^\pi \mathbf{n}(t_0) \cdot \mathbf{n}(t_0 + t) dt + \int_0^\pi \mathbf{n}(t_0) \cdot \mathbf{n}(t_0 - t) dt > 0,$$

so for all t_0 , $\mathbf{n}(t_0) \in \{\mathbf{n} \in \mathbb{S}^{d-1} \mid \mathbf{n} \cdot \mathbf{M} > 0\}$. \square

Now the statements (5.12) and (5.13) on the non-local-controllability of $(\mathbb{M}, d_{\mathcal{F}_0^+})$ are shown in two steps.

Step 1: we show in the case of a constant cost function $\mathcal{C}_2 = \delta$ one has

$$\limsup_{\mathbf{p}' \rightarrow \mathbf{p}} d_{\mathcal{F}_0^+}(\mathbf{p}, \mathbf{p}') \leq 2\pi\delta,$$

for any $\mathbf{p} \in \mathbb{M}$. Indeed, one can design an admissible curve in $(\mathbb{M}, \mathcal{F}_0^+)$ as the concatenation of an in-place rotation, a straight line, and an in-place rotation. The length of the straight line is $\mathcal{O}(\|\mathbf{p}' - \mathbf{p}\|)$ and vanishes when $\mathbf{p}' \rightarrow \mathbf{p}$, and the in-place rotations each have maximum cost $\pi\delta$.

Step 2: we prove the lower bound

$$\lim_{\mu \downarrow 0} d_{\mathcal{F}_0^+}((\mathbf{x}, \mathbf{n}), (\mathbf{x} - \mu\mathbf{n}, \mathbf{n})) \geq 2\pi\delta,$$

for any $(\mathbf{x}, \mathbf{n}) \in \mathbb{M}$. This and the above established upper bound implies the required result. As $\mathcal{C}_1, \mathcal{C}_2 \geq \delta$, we can restrict ourselves to the case of uniform cost $\mathcal{C}_1 = \mathcal{C}_2 = \delta = 1$ and just show equality (5.13), as the estimate (5.12) follows by scaling with δ .

Consider a Lipschitz regular path $\gamma(t) = (\mathbf{x}(t), \mathbf{n}(t))$, with $\dot{\mathbf{x}} \propto \mathbf{n}$ and $\dot{\mathbf{x}} \cdot \mathbf{n} \geq 0$, from (\mathbf{x}, \mathbf{n}) to $(\mathbf{x} - \mu\mathbf{n}, \mathbf{n})$. Then

$$\mathbf{0} = \mu\mathbf{n} + \int_0^1 \dot{\mathbf{x}}(t) dt = \mu\mathbf{n}(0) + \int_0^1 \|\dot{\mathbf{x}}(t)\| \mathbf{n}(t) dt,$$

so $\mathbf{0} \in \text{Hull}\{\mathbf{n}(t); 0 \leq t \leq 1\}$. Let $\mathbf{m} : [0, 1] \rightarrow \mathbb{S}^{d-1}$ be a constant speed parametrization of \mathbf{n} . Let $\tilde{\mathbf{m}} : \mathbb{R} \rightarrow \mathbb{S}^{d-1}$ be defined by $\tilde{\mathbf{m}}(2\pi t) = \mathbf{m}(t)$ for all $t \in [0, 2\pi]$, and extended by 2π -periodicity. If $\tilde{\mathbf{m}}(\cdot)$ were strictly 1-Lipschitz then by Lemma 5.3.1 we would get $\mathbf{0} \notin \text{Hull}\{\tilde{\mathbf{m}}(t) \mid t \in [0, 2\pi]\} = \text{Hull}\{\mathbf{n}(t) \mid t \in [0, 1]\}$ and a contradiction. Hence there exists a $t_0 \in \mathbb{R}$ such that $\|\dot{\tilde{\mathbf{m}}}(t_0)\| \geq 1$ and via the constant speed parametrization assumption we get the required coercivity:

$$1 \leq \|\dot{\tilde{\mathbf{m}}}(t_0)\| = \frac{1}{2\pi} \int_0^1 \|\dot{\mathbf{n}}(t)\| dt \Rightarrow \int_0^1 \mathcal{F}_0^+(\gamma(t), \dot{\gamma}(t)) dt \geq \int_0^1 \mathcal{C}_2(\gamma(t)) \|\dot{\mathbf{n}}(t)\| dt \geq 2\pi\delta.$$

To prove local controllability of the model $(\mathbb{M}, d_{\mathcal{F}_0})$, we apply the logarithmic approximation for weighted sub-coercive operators on Lie groups, cf. [tER98] applied to the Lie group $SE(d) = \mathbb{R}^d \rtimes SO(d)$, in which the space of positions and orientations is placed via a Lie group quotient $SE(d)/(\{0\} \times SO(d-1))$. One obtains a sharp estimate (for specific sharp estimates for $d = 3$, in the context of heat-kernels estimation, see Chapter 4), where the weights of allowable (horizontal) vector fields is 1, whereas the remaining spatial vector fields orthogonal to $\mathbf{n} \cdot \nabla_{\mathbb{R}^d}$ get weight 2, as they follow by a single commutator of allowable vector fields, see e.g. [DGDHM16, DGDHS14]. Relaxing all spatial weights to 2 and continuity of costs $\mathcal{C}_1, \mathcal{C}_2$, yields (5.14). \square

Remark 5.3.2. *In view of the above one might expect that the point $(\mathbf{x} - \mu\mathbf{n}, \mathbf{n})$ is reached by a geodesic that consists of 1. an in-place rotation by π , 2. a straight line, 3. an in-place rotation by π . However, this is not the case as can be observed in the very lower left corner in Fig. 5.6, where the two minimizing red curves show a very different behavior. This is explained by the next lemma.*

Lemma 5.3.3. *Let $\mu > 0$, and $\mathcal{C}_1 = \mathcal{C}_2 = \delta$. Let \mathbf{R}_θ denote the (counter-clockwise) rotation matrix about the origin by angle θ . The endpoint $(\mathbf{x} - \mu\mathbf{n}, \mathbf{n})$ for each $\mu \geq 0$ is a Maxwell point w.r.t. (\mathbf{x}, \mathbf{n}) , since there are two minimizing geodesics in $(\mathbb{M}, d_{\mathcal{F}_0^+})$ that are a concatenation*

1. an in-place rotation from (\mathbf{x}, \mathbf{n}) to $(\mathbf{x}, \mathbf{R}_{\pm\frac{\pi}{2}}\mathbf{n})$,
2. a full U-curve, see [MS10], departing from and ending in a cusp from $(\mathbf{x}, \mathbf{R}_{\pm\frac{\pi}{2}}\mathbf{n})$ to $(\mathbf{x} - \mu\mathbf{n}, \mathbf{R}_{\mp\frac{\pi}{2}}\mathbf{n})$,
3. an in-place rotation from $(\mathbf{x} - \mu\mathbf{n}, \mathbf{R}_{\mp\frac{\pi}{2}}\mathbf{n})$ to $(\mathbf{x} - \mu\mathbf{n}, \mathbf{n})$.

We have the limit $\lim_{\mu \downarrow 0} d_{\mathcal{F}_0^+}((\mathbf{x}, \mathbf{n}), (\mathbf{x} - \mu\mathbf{n}, \mathbf{n})) = 2\pi\delta$.

Proof. See [DMMP16]. \square

Remark 5.3.4. *Consider the case $d = 2$, $\mathcal{C}_1 = \mathcal{C}_2 = \delta$, and source point $\mathbf{p}_S = (\mathbf{x}, \mathbf{n}) = \mathbf{e} = (0, 0, \theta = 0)$. The end-points $(\mathbf{x} - \mu\mathbf{n}, \mathbf{n}) = (-\mu, 0, 0)$, with $\mu > 0$ sufficiently small, are 1st Maxwell-points in $(\mathbb{M}, d_{\mathcal{F}_0^+})$ where geodesically equidistant wavefronts departing from the source point collide for the first time, see Fig. 5.7C. The distance mapping $d_{\mathcal{F}_0^+}^+(\mathbf{p}_S, \cdot)$ is not continuous, but the asymmetric distance spheres*

$$\mathcal{S}_R := \{\mathbf{p} \in \mathbb{M} \mid d_{\mathcal{F}_0^+}^+(\mathbf{p}_S, \mathbf{p}) = R\}$$

are connected and compact, and they collide at $R = 2\pi$ in such a way that the origin \mathbf{p}_s becomes an interior point in the asymmetric balls of radius $R > 2\pi$.

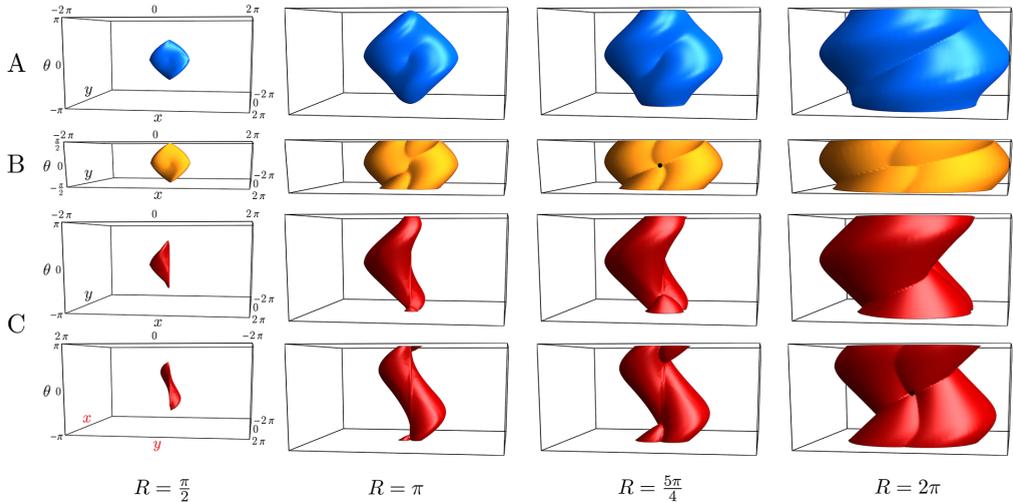


Figure 5.7: The development of spheres centered around $\mathbf{e} = (0, 0, 0)$ with increasing radius R . **A:** the normal SR spheres on \mathbb{M} given by $\{\mathbf{p} \in \mathbb{M} \mid d_{\mathcal{F}_0}(\mathbf{p}, \mathbf{e}) = R\}$ where the folds reflect the 1st Maxwell sets [BDMS15, Sac11]. **B:** the SR spheres with identification of antipodal points given by $\{\mathbf{p} \in \mathbb{M} \mid \min\{d_{\mathcal{F}_0}(\mathbf{p}, \mathbf{e}), d_{\mathcal{F}_0}(\mathbf{p} + (0, 0, \pi), \mathbf{e})\} = R\}$ with additional folds (1st Maxwell sets) due to π -symmetry. **C:** the asymmetric Finsler norm spheres given by $\{\mathbf{p} \in \mathbb{M} \mid d_{\mathcal{F}_0^+}(\mathbf{p}, \mathbf{e}) = R\}$ visualized from two perspectives with extra folds (1st Maxwell sets) at the back $(-\mu, 0, 0)$. The black dots indicate points with two folds. In the case of B, this is a Maxwell-point with 4 geodesics merging. In the case of C, this is just the origin itself reached from behind at $R = 2\pi$, recall Lemma 5.3.3. Although not depicted here, if the radius $R > 2\pi$ the origin becomes an interior point of the corresponding ball.

5.4 Cusps and keypoints: proof of Theorem 5.2.10

In this section we provide a proof of Theorem 5.2.10 on the occurrence of cusps and keypoints. For the uniform cost case $\mathcal{C}_1 = \mathcal{C}_2 = 1$ for $d = 2$, our curve-optimization problem (5.1) $(\mathbb{M}, d_{\mathcal{F}_0})$ in consideration, boils down to a standard left-invariant curve optimization in the roto-translation group $SE(2) = \mathbb{R}^2 \times SO(2)$. As we will apply tools from previous works [DBRS13, BDRS14, BCR10, Sac11], we will make use of the following notations for expansion⁵ of velocity and momentum in the left-invariant (co)-frame:

$$\begin{cases} \mathcal{A}_1 := \cos \theta \partial_x + \sin \theta \partial_y, \\ \mathcal{A}_2 := -\sin \theta \partial_x + \cos \theta \partial_y, \\ \mathcal{A}_3 := \partial_\theta, \end{cases} \quad \begin{cases} \omega^1 := \cos \theta dx + \sin \theta dy, \\ \omega^2 := -\sin \theta dx + \cos \theta dy, \\ \omega^3 := d\theta, \end{cases} \quad (5.32)$$

$$\dot{\gamma}(t) = \sum_{i=1}^3 u^i(t) \mathcal{A}_i|_{\gamma(t)} \in T_{\gamma(t)}(\mathbb{M}), \quad \hat{\mathbf{p}}(t) = \sum_{i=1}^3 \hat{p}_i(t) \omega^i|_{\gamma(t)} \in T_{\gamma(t)}^*(\mathbb{M}),$$

⁵Note that we use upper-indices for the control's (velocity components) as they are contra-variant.

where the indexing of the left-invariant frame is different here, in order to stick to the ordering (x, y, θ) applied in this article. Note that for the case $\varepsilon = 0$ admissible smooth curves γ in $(\mathbb{M}, d_{\mathcal{F}_0})$ satisfy the horizontality constraint $\dot{\gamma}(t) \in \text{Span}\{\mathcal{A}_1|_{\gamma(t)}, \mathcal{A}_3|_{\gamma(t)}\}$.

Proof of the statements regarding cusps:

- We can describe our curve optimization problem (5.1) using a Hamiltonian formalism, with Hamiltonian $H(\hat{\mathbf{p}}) = \frac{1}{2}(\hat{p}_1^2 + \hat{p}_3^2) = \frac{1}{2}$ [MS10]. By Pontryagin's Maximum Principle, geodesics adhere to the following Hamilton equations:

$$\begin{cases} \dot{p}_1 = u^1 = \hat{p}_1, \\ \dot{p}_2 = u^2 = 0, \\ \dot{p}_3 = u^3 = \hat{p}_3, \end{cases}, \quad \begin{cases} \frac{d\hat{p}_1}{dt} = \hat{p}_2\hat{p}_3, \\ \frac{d\hat{p}_2}{dt} = -\hat{p}_1\hat{p}_3, \\ \frac{d\hat{p}_3}{dt} = -\hat{p}_1\hat{p}_2. \end{cases} \quad (5.33)$$

For fixed initial momentum $\hat{\mathbf{p}}(0)$, this uniquely determines a SR geodesic. Moreover, SR geodesics are contained within the (co-adjoint) orbits

$$(\hat{p}_1(t))^2 + (\hat{p}_2(t))^2 = (\hat{p}_1(0))^2 + (\hat{p}_2(0))^2. \quad (5.34)$$

The parameter t in the system (5.33) is SR arc length, but by reparametrizing (possible as long as u^1 does not change sign) to spatial arc length parameter s , with $\frac{ds}{dt} = \hat{p}_1$, we get a partially linear system. Combining (5.33) and (5.34), we find orbits in the (hyperbolic) phase portrait induced by

$$\begin{cases} \hat{p}'_2(s) = -\hat{p}_3 \\ \hat{p}'_3(s) = -\hat{p}_2 \end{cases} \implies \begin{cases} \hat{p}_2(s) = \hat{p}_2(0) \cosh s - \hat{p}_3(0) \sinh s \\ \hat{p}_3(s) = -\hat{p}_2(0) \sinh s + \hat{p}_3(0) \cosh s. \end{cases}$$

Hence $|\hat{p}_3(s)| = 1$ always has a solution for some finite (possibly negative) s , except when $\hat{p}^2(0) = \hat{p}_3(0) = 0$, in which case the solutions are straight lines. Preservation of the Hamiltonian then implies $\hat{p}_1(s) = u^1(s) = \tilde{u}(s) = 0$. We conclude that every SR geodesic (with unconstrained time $t \in \mathbb{R}$) in $(\mathbb{M}, d_{\mathcal{F}_0})$ which is not a straight line admits a cusp.

- We now consider $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$, $\varepsilon > 0$. To have a cusp, we need $\hat{p}_1(t) = \hat{p}_2(t) = 0$ for some $t \in \mathbb{R}$. The co-adjoint orbit condition (5.34) then implies that $\hat{p}_1(t) = \hat{p}_2(t) = 0$ for all t , corresponding to a vertical geodesic that has purely angular momentum and no cusp. The same argument holds for $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$. In $(\mathbb{M}, d_{\mathcal{F}_0^+})$ we have the condition that $u^1 \geq 0$, hence by definition it can never switch sign and all geodesics are cusplless.

Proof of the statements regarding keypoints:

- For the cases $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$ and $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$ with $\varepsilon > 0$ we can use the same line of arguments as above. Also here both spatial controls have to vanish, resulting in vertical geodesics. The spatial projection of such curves is a single keypoint. For $(\mathbb{M}, d_{\mathcal{F}_0})$ we rely on the result that SR geodesics are analytical, and therefore if the control $u^1(t) = 0$ for some open time interval (t_0, t_1) , then $u^1(t) = 0$ for all $t \in \mathbb{R}$, again corresponding to purely angular motion.
- Geodesics in $(\mathbb{M}, d_{\mathcal{F}_0^+})$ can have keypoints only at the boundaries. Suppose a geodesic $\gamma : [0, 1] \rightarrow \mathbb{M}$ in $(\mathbb{M}, d_{\mathcal{F}_0^+})$ has an internal keypoint, with a corner of angle $\delta > 0$, at internal time $T_1 \in (0, 1)$. Then one can create a local shortcut with a straight line segment connecting two sufficiently close points before and after the corner with two in-place rotations whose angles add up to δ . With a suitable mollifier this shortcut can be approximated by a curve in Γ . For details see similar arguments in [BDRS14].

Next we explain the cases A), B) and C), where we fix initial point $\gamma(0) = \mathbf{e} = (0, 0, 0)$.

- A) Suppose that the endpoint $\mathbf{p} = (x, y, \theta) \in \overline{\mathfrak{R}}$ and $x \geq 0$. Then \mathbf{p} can already be reached by a geodesic in $(\mathbb{M}, d_{\mathcal{F}_0})$ and the positivity constraint (i.e. no reverse gear), which can only increase length, becomes obsolete.
- B) Now suppose the endpoint $\mathbf{p} = (x, y, \theta)$ lays in the half-space $x < 0$. Then by the half-space property of geodesics in $(\mathbb{M}, d_{\mathcal{F}_0})$, cf. [DBRS13, Thm.7], the geodesic in $(\mathbb{M}, d_{\mathcal{F}_0^+})$ must have a keypoint. By the preceding keypoints can only be located at the boundaries. If it takes place at the endpoint only, then still the constraint $x < 0$ is not satisfied, thereby it must take place at the origin.
- C) In those cases the endpoint \mathbf{p} lays outside the connected cone of reachable angles, which are by [DBRS13, Thm.9] bounded (for those endpoints) by geodesics ending in a cusp (so not endpoints of geodesics starting at a cusp). So for those points, minimizing geodesics will first move by an in-place rotation (along a spherical geodesic) until it hits the cusp surface $\partial\mathfrak{R}$, after which it is traced back to the origin by a regular geodesic with strictly positive spatial control inside the volume \mathfrak{R} .

5.5 Eikonal equations and backtracking: proof of Prop. 5.2.13, Corr. 5.2.14 and Thm. 5.2.15

First we shall prove Proposition 5.2.13, regarding the duals of \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$, and Corollary 5.2.14, providing explicit expressions for the corresponding eikonal equations. To

this end we need a basic lemma on computing dual norms on \mathbb{R}^n , where later we will set $n = 2d - 1 = \dim(\mathbb{M})$.

Lemma 5.5.1. *Let $\mathbf{w} \in \mathbb{R}^n$ and let $\mathbf{M} \in \mathbb{R}^{n \times n}$ be symmetric, positive definite. Define the norm $F_{\mathbf{M}, \mathbf{w}} : \mathbb{R}^n \rightarrow \mathbb{R}^+$ by*

$$F_{\mathbf{M}, \mathbf{w}}(\mathbf{v}) = \sqrt{(\mathbf{M}\mathbf{v}, \mathbf{v}) + (\mathbf{w}, \mathbf{v})_-^2}.$$

Then its dual norm $F_{\mathbf{M}, \mathbf{w}}^* : (\mathbb{R}^n)^* \rightarrow \mathbb{R}^+$ equals

$$F_{\mathbf{M}, \mathbf{w}}^*(\hat{\mathbf{v}}) = \sqrt{(\hat{\mathbf{v}}, \hat{\mathbf{M}}\hat{\mathbf{v}}) + (\hat{\mathbf{v}}, \hat{\mathbf{w}})_+^2}, \quad (5.35)$$

with $\hat{\mathbf{M}} = (\mathbf{M} + \mathbf{w} \otimes \mathbf{w})^{-1}$ and $\hat{\mathbf{w}} = \frac{\mathbf{M}^{-1}\mathbf{w}}{\sqrt{1 + (\mathbf{w}, \mathbf{M}^{-1}\mathbf{w})}}$.

Proof. For $n = 1$ the result is readily verified, and for $\mathbf{w} = \mathbf{0}$ the result is classical. We next turn to the special case $\mathbf{M} = \text{Id}$, and $\mathbf{w} = (w_1, \mathbf{0}_{\mathbb{R}^{n-1}})$ is zero except maybe for its first coordinate w_1 . Thus for any $\mathbf{v} = (v_1, \mathbf{v}_2) \in \mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$ one has the splitting

$$F_{\mathbf{M}, \mathbf{w}}(v_1, \mathbf{v}_2)^2 = (|v_1|^2 + (w_1 v_1)_-^2) + \|\mathbf{v}_2\|^2 := F_1(v_1)^2 + F_2(\mathbf{v}_2)^2. \quad (5.36)$$

Using the compatibility of norm duality with such splittings, and the special cases $n = 1$ and $\mathbf{w} = \mathbf{0}$ mentioned above, we obtain

$$(F_{\mathbf{M}, \mathbf{w}}^*(\hat{v}_1, \hat{\mathbf{v}}_2))^2 = (F_1^*(\hat{v}_1))^2 + (F_2^*(\hat{\mathbf{v}}_2))^2 = \frac{|\hat{v}_1|^2 + (w_1 \hat{v}_1)_+^2}{1 + |w_1|^2} + \|\hat{\mathbf{v}}_2\|^2,$$

which is exactly of the form (5.35). The general case for arbitrary \mathbf{w} and symmetric positive definite \mathbf{M} follows from affine invariance. Indeed let \mathbf{A} be an invertible $n \times n$ matrix, and let $\mathbf{M}' = \mathbf{A}^T \mathbf{M} \mathbf{A}$ and $\mathbf{w}' = \mathbf{A}^T \mathbf{w}$. Let $F = F_{\mathbf{M}, \mathbf{w}}$ and $F' = F_{\mathbf{M}', \mathbf{w}'}$, so that $F'(\mathbf{v}) = F(\mathbf{A}\mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^n$. Let F^* , $\hat{\mathbf{M}}$, $\hat{\mathbf{w}}$, and F'^* , $\hat{\mathbf{M}}'$, $\hat{\mathbf{w}}'$, be respectively the dual norms and the matrices defined by the explicit formulas above. Then denoting $\mathbf{B} := (\mathbf{A}^T)^{-1}$ one has by the definition of dual norms that $F'^*(\hat{\mathbf{v}}) = F^*(\mathbf{B}\hat{\mathbf{v}})$ for all $\hat{\mathbf{v}} \in \mathbb{R}^n$, and by the explicit formulas $\hat{\mathbf{M}}' = \mathbf{B}^T \hat{\mathbf{M}} \mathbf{B}$, $\mathbf{w}' = \mathbf{B}^T \mathbf{w}$. Thus, $F^* = F_{\mathbf{M}, \mathbf{w}}^*$ holds if and only if $F'^* = F_{\mathbf{M}', \mathbf{w}'}^*$. Since for any \mathbf{M}, \mathbf{w} , there exists a linear change of variables \mathbf{A} such that $\mathbf{M}' = \text{Id}$ and \mathbf{w}' is zero except maybe for its first coordinate, the proof is complete. \square

Now Proposition 5.2.13 follows from Lemma 5.5.1 by writing out the dual norm, using for each $\mathbf{p} \in \mathbb{M}$:

$$\begin{aligned} \mathbf{M}_{\mathbf{p}} &= (\mathcal{C}_1(\mathbf{p}))^2 (\mathbf{D}_{\mathbf{n}}^\varepsilon)^{-1} \oplus (\mathcal{C}_2(\mathbf{p}))^2 I_d \quad \text{and} \\ \mathbf{w}_{\mathbf{p}} &= \begin{cases} \mathcal{C}_1(\mathbf{p}) \sqrt{\varepsilon^{-2} - 1} (\mathbf{n}, \mathbf{0}), & \text{for } \mathcal{F}_{\varepsilon^+}, \\ \mathbf{0}, & \text{for } \mathcal{F}_{\varepsilon}, \end{cases} \end{aligned} \quad (5.37)$$

with $\mathbf{D}_{\mathbf{n}}^\varepsilon$ as in (5.26). Corollary 5.2.14 then follows by setting the momentum covector $\hat{\mathbf{p}} = dU(\mathbf{p})$ equal to the derivative of the value function evaluated at \mathbf{p} .

Now that we have derived the eikonal equations, we obtain the backtracking Theorem 5.2.15 by Proposition B.2.2 in App. B.2, which shows us that level sets of solutions of the eikonal equations are geodesically equidistant surfaces and that geodesics are found by an intrinsic gradient descent.

However, to obtain the explicit backtracking formulas we differentiate the Hamiltonian, rather than the dual metric, which is equivalent thanks to (B.6) (in Remark B.2.3 in App. B.2). We focus below on the model $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$ without reverse gear, since the other case is similar. Let $\mathbf{p} \in \mathbb{M}$, let $F := \mathcal{F}_\varepsilon^+(\mathbf{p}, \cdot)$, and let $\hat{\mathbf{p}} = (\hat{\mathbf{x}}, \hat{\mathbf{n}}) \in T_{\mathbf{p}}^*(\mathbb{M})$. Then differentiating w.r.t. $\hat{\mathbf{n}}$ we obtain

$$d_{\hat{\mathbf{n}}}F^*(\hat{\mathbf{x}}, \hat{\mathbf{n}})^2 = \mathcal{C}_2(\mathbf{p})^{-2} d_{\hat{\mathbf{n}}}\|\hat{\mathbf{n}}\|^2 = 2\mathcal{C}_2(\mathbf{p})^{-2}\hat{\mathbf{n}},$$

where $\|\cdot\|$ is the Riemannian metric induced by the embedding $S^{d-1} \subset \mathbb{R}^d$. Differentiating w.r.t. $\hat{\mathbf{x}}$ we obtain

$$d_{\hat{\mathbf{x}}}F^*(\hat{\mathbf{x}}, \hat{\mathbf{n}})^2 = \mathcal{C}_1(\mathbf{p})^{-2} d_{\hat{\mathbf{x}}}(\hat{\mathbf{x}} \cdot \mathbf{D}_{\mathbf{n}}^\varepsilon \hat{\mathbf{x}} - (1 - \varepsilon^2)(\hat{\mathbf{x}} \cdot \mathbf{n})_-^2) \quad (5.38)$$

$$= 2\mathcal{C}_1(\mathbf{p})^{-2} \begin{cases} \mathbf{D}_{\mathbf{n}}^\varepsilon \hat{\mathbf{x}} & \text{if } \hat{\mathbf{x}} \cdot \mathbf{n} \geq 0, \\ \varepsilon^2 \text{Id } \hat{\mathbf{x}} & \text{if } \hat{\mathbf{x}} \cdot \mathbf{n} \leq 0. \end{cases} \quad (5.39)$$

The announced result (5.30), which is equivalent to its more concise abstract form (5.27), follows by choosing $\hat{\mathbf{x}} := \nabla_{\mathbb{R}^d}U(\gamma(t))$ and $\hat{\mathbf{n}} := \nabla_{S^{d-1}}U(\gamma(t))$ and a basic re-scaling $[0, L] \in t \mapsto t/L \in [0, 1]$. \square

Remark 5.5.2. *The computation of the dual norms can be simplified by expressing velocity (entering the Finsler metric) and momentum (entering the dual metric) in a (left-invariant) local, orthogonal, moving frame of reference, attached to the point $\mathbf{p} = (\mathbf{x}, \mathbf{n}) \in \mathbb{M}$:*

$$\dot{\mathbf{p}} = \sum_{i=1}^{2d-1} u^i \mathcal{A}_i|_{\mathbf{p}}, \quad \hat{\mathbf{p}} = \sum_{i=1}^{2d-1} \hat{p}_i \omega^i|_{\mathbf{p}} \quad (5.40)$$

where a moving frame of reference is chosen such that

$$u^d = \tilde{u} = \mathbf{n} \cdot \dot{\mathbf{x}}, \quad \sum_{i=1}^{d-1} (u^i)^2 = \|\dot{\mathbf{x}}\|^2 - (\mathbf{n} \cdot \dot{\mathbf{x}})^2, \quad \sum_{i=1}^{d-1} (u^{d+i})^2 = \|\hat{\mathbf{n}}\|^2,$$

inducing a corresponding dual frame $\{\omega^i|_{\mathbf{p}}\}$ via

$$\langle \omega^i|_{\mathbf{p}}, \mathcal{A}_j|_{\mathbf{p}} \rangle = \delta_j^i, \quad \text{for all } i, j = 1, \dots, 2d-1. \quad (5.41)$$

W.r.t. the left-invariant frame the matrices $\mathbf{D}_{\mathbf{n}}^\varepsilon$, $M_{\mathbf{p}}$ as in (5.37) and $\hat{M}_{\mathbf{p}}$ all become diagonal matrices, and the dual can be computed straightforwardly. Furthermore, in this formulation we can see from the expression for the dual $(\mathcal{F}_0^+)^*$, i.e. in the limit $\varepsilon \downarrow 0$, that the positive spatial control u^d constraint results in a positive momentum \hat{p}_d constraint:

$$(\mathcal{F}_0^+)^*(\mathbf{p}, \hat{\mathbf{p}}) = \sqrt{\frac{(\hat{p}_d)_+^2}{\mathcal{C}_1^2(\mathbf{p})} + \frac{1}{\mathcal{C}_2^2(\mathbf{p})} \sum_{i=d+1}^{2d-1} (\hat{p}_i)^2}. \quad (5.42)$$

Therefore the eikonal equation in the positive control model $(\mathbb{M}, d_{\mathcal{F}_0^+})$ is simply given by

$$\sqrt{\frac{\|\nabla_{S^{d-1}} U(\mathbf{p})\|^2}{\mathcal{C}_2^2(\mathbf{p})} + \frac{((\mathbf{n} \cdot \nabla_{\mathbb{R}^d} U(\mathbf{p}))_+)^2}{\mathcal{C}_1^2(\mathbf{p})}} = 1 \quad (5.43)$$

5.6 Discretization of the Eikonal PDEs

5.6.1 Causal operators and the fast marching algorithm

The fast marching algorithm is an efficient numerical method [Tsi95] for numerically solving the static first order Hamilton-Jacobi-Bellman (or simply eikonal) PDE (5.5) which characterizes the distance map U to a fixed source point \mathbf{p}_S . Fast marching is tightly connected with Dijkstra's algorithm on graphs, and in particular it shares the $\mathcal{O}(KN \ln N)$ complexity, where $N = \#(X)$ is the cardinality of the discrete domain $X \subset \mathbb{M}$, $X \ni \mathbf{p}_S$, and K is the average number of neighbors for each point. Both fast marching and Dijkstra's algorithms can be regarded as specialized solvers of non-linear fixed point systems of equations $\Lambda u = u$, where the unknown $u \in \mathbb{R}^X$ is a discrete map representing the front arrival times, which rely on the a-priori assumption that the operator $\Lambda : \mathbb{R}^X \rightarrow \mathbb{R}^X$ is *causal* (and monotone, but this second assumption is not discussed here). Causality informally means that the estimated front arrival time $\Lambda u(\mathbf{p})$ at a point $\mathbf{p} \in X$ depends on the given arrival times $u(\mathbf{q})$, $\mathbf{q} \in X$, prior to $\Lambda u(\mathbf{p})$, but not on the simultaneous or the future ones. Formally, one requires that for any $u, v \in \mathbb{R}^X$, $t \in \mathbf{R}$:

$$\text{If } u^{<t} = v^{<t} \text{ then } (\Lambda u)^{\leq t} = (\Lambda v)^{\leq t}, \quad \text{where } u^{<t}(\mathbf{p}) := \begin{cases} u(\mathbf{p}) & \text{if } u(\mathbf{p}) < t, \\ +\infty & \text{otherwise,} \end{cases} \quad (5.44)$$

and $v^{<t}$, $(\Lambda u)^{\leq t}$ and $(\Lambda v)^{\leq t}$ are defined similarly.

A semi-Lagrangian scheme. We implemented two discretizations of the eikonal equation (5.5) which benefit from the causality property. The first one is a semi-Lagrangian scheme, inspired by *Bellman's optimality principle* which informally states

that any sub-policy of an optimal policy is an optimal policy. Formally, let \mathcal{F} be a Finsler metric, and let $U := d_{\mathcal{F}}(\cdot, \mathbf{p}_S)$ be defined as the distance to a given source point \mathbf{p}_S . Then for any $\mathbf{p} \in \mathbb{M}$ and any neighborhood V of \mathbf{p} not containing \mathbf{p}_S one has the property

$$U(\mathbf{p}) := \min_{\mathbf{q} \in \partial V} d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) + U(\mathbf{q}). \quad (5.45)$$

In the spirit of [Tsi95, SV01] we discretize (5.45) by introducing for each interior $\mathbf{p} \in X \setminus \{\mathbf{p}_S\}$ a small polygonal neighborhood $V(\mathbf{p})$, which vertices belong to the discrete point set X . The nonlinear operator Λ is defined as

$$\Lambda u(\mathbf{p}) := \min_{\substack{\{\mathbf{q}_1, \dots, \mathbf{q}_n\} \\ \text{facet of } \partial V(\mathbf{p})}} \min_{\xi \in \Xi} \mathcal{F} \left(\mathbf{p}, \sum_{i=1}^n \xi_i \mathbf{q}_i - \mathbf{p} \right) + \sum_{i=1}^n \xi_i u(\mathbf{q}_i), \quad (5.46)$$

where $\Xi = \{\xi \in \mathbb{R}_+^n; \sum_{i=1}^n \xi_i = 1\}$. In other words, the boundary point $\mathbf{q} \in \partial V(\mathbf{p})$ in (5.45) is represented in (5.46) by the barycentric sum $\mathbf{q} = \sum_{i=1}^n \xi_i \mathbf{q}_i$, the distance $d_{\mathcal{F}}(\mathbf{p}, \mathbf{q})$ is approximated with the norm $\mathcal{F}(\mathbf{p}, \mathbf{q} - \mathbf{p})$, and the value $U(\mathbf{q})$ is approximated with the interpolation $\sum_{i=1}^n \xi_i u(\mathbf{q}_i)$.

We refer to [SV01, Vla06] for proofs of convergence, and for the following essential property: the operator (5.46) obeys the causality property (5.44) iff the chosen stencil $V(\mathbf{p})$ obeys the following generalized acuteness property: for any \mathbf{q}, \mathbf{q}' in a common facet of $V(\mathbf{p})$, one has

$$\langle d_{\mathbf{p}} \mathcal{F}(\mathbf{p}, \mathbf{q} - \mathbf{p}), \mathbf{q}' - \mathbf{p} \rangle \geq 0.$$

For the construction of such stencils $V(\mathbf{p})$, $\mathbf{p} \in X$, we rely on the previous works [Mir14, Mir13] and on the following observation: the metrics $\mathcal{F}_{\varepsilon}$ and $\mathcal{F}_{\varepsilon}^+$ associated to the Reeds-Shepp car models can be decomposed as

$$\mathcal{F}(\mathbf{p}, (\dot{\mathbf{x}}, \dot{\mathbf{n}}))^2 = \mathcal{F}_1(\mathbf{p}, \dot{\mathbf{x}})^2 + \mathcal{F}_2(\mathbf{p}, \dot{\mathbf{n}})^2, \quad (5.47)$$

which allows to build the stencils $V(\mathbf{p})$ for \mathcal{F} by combining, as discussed in [Mir14, p. 9], some lower dimensional stencils $V_1(\mathbf{p})$ and $V_2(\mathbf{p})$ built independently for for the spatial $\mathbf{x} \in \mathbb{R}^d$ and spherical $\mathbf{n} \in S^{d-1}$ variables.

We discretize S^1 uniformly, with the standard choice of stencil. We discretize S^2 by refining uniformly the faces of an icosahedron and projecting their vertices onto the sphere (as performed by the Mathematica[®] Geodesate function). The resulting triangulation only features acute interior angles, in the classical Euclidean sense, and thus provides adequate stencils since in our applications $\mathcal{F}_2(\mathbf{p}, \dot{\mathbf{n}}) = \mathcal{C}_2(\mathbf{p}) \|\dot{\mathbf{n}}\|$ is proportional to the Euclidean norm, see Fig. 5.8. We typically use 60 discretization points for S^1 , and from 200 to 2000 points for S^2 .

We discretize \mathbb{R}^d using the Cartesian grid $h\mathbb{Z}^d$, where $h > 0$ is the discretization scale. The norm $\mathcal{F}_{\varepsilon,1}(\mathbf{p}, \dot{\mathbf{x}}) = \mathcal{C}_1(\mathbf{p}) \sqrt{\dot{\mathbf{x}}^T (\mathbf{D}_{\mathbf{n}}^{\varepsilon})^{-1} \dot{\mathbf{x}}}$, recall the notation in (5.47), that

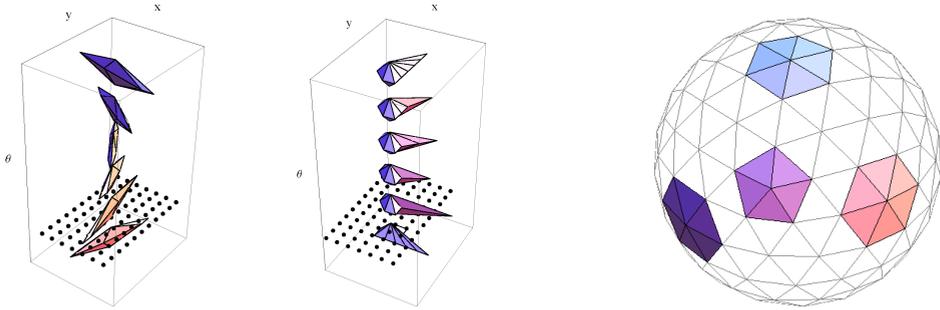


Figure 5.8: Left: Stencil used for the metric \mathcal{F}_ε on $\mathbb{R}^2 \times S^1$, $\varepsilon = 0.1$, obeying the generalized acuteness property required for the Bellman type discretization (5.46). See also the control sets in Fig. 5.2. Center: likewise with $\mathcal{F}_\varepsilon^+$, $\varepsilon = 0.1$. Right: Coarse discretization of S^2 with 162 vertices, used in some experiments posed on $\mathbb{R}^3 \times S^2$. Some acute stencils (in the classical Euclidean sense) shown in color.

is induced by the approximate Finsler metric \mathcal{F}_ε on the physical variables in \mathbb{R}^d , is of Riemannian type and is strongly anisotropic. In dimension $d \leq 3$, this is the adequate setting for the adaptive stencils of [Mir14], built using discrete geometry tools known as lattice basis reduction. The norm $\mathcal{F}_{\varepsilon,1}^+(\mathbf{p}, \dot{\mathbf{x}}) = \mathcal{C}_1(\mathbf{p}) \sqrt{\dot{\mathbf{x}}^T (\mathbf{D}_\mathbf{n}^\varepsilon)^{-1} \dot{\mathbf{x}} + (\varepsilon^{-2} - 1) (\mathbf{n}, \mathbf{x})_-^2}$ induced by $\mathcal{F}_\varepsilon^+$ on \mathbb{R}^d is Finslerian (i.e. non-Riemannian) and strongly anisotropic. In dimension $d = 2$, this is the adequate setting for the adaptive stencils of [Mir13], built using an arithmetic object known as the Stern-Brocot tree.

Direct approximation of the Hamiltonian. A new approach, not semi-Lagrangian, had to be developed for the Finsler metric $\mathcal{F}_\varepsilon^+$ in dimension $d = 3$ due to our failure to construct viable (i.e. with a reasonably small number of reasonably small vertices) stencils obeying the generalized acuteness property in this case, see Fig. 5.9. For manuscript size reasons, we only describe it informally, and postpone proofs of convergence for future work.

Let $\mathbf{n} \in S^2$ and let $\varepsilon > 0$ be fixed. Then one can find *non-negative* weights and *integral* vectors $(\rho_i, \mathbf{w}_i) \in (\mathbf{R}_+ \times \mathbb{Z}^3)^6$, such that for all $\mathbf{v} \in \mathbb{R}^3$

$$\sum_{1 \leq i \leq 6} \rho_i (\mathbf{w}_i \cdot \mathbf{v})^2 = (\mathbf{n} \cdot \mathbf{v})^2 + \varepsilon^2 \|\mathbf{n} \times \mathbf{v}\|^2. \quad (5.48)$$

A simple and efficient construction of $(\rho_i, \mathbf{w}_i)_{i=1}^6$, relying on the concept of obtuse superbase of a lattice, is in [FM14] described and used to discretize anisotropic diffusion PDEs. One may furthermore assume that $(\mathbf{n}, \mathbf{w}_i) \geq 0$ for all $1 \leq i \leq 6$, up to replacing

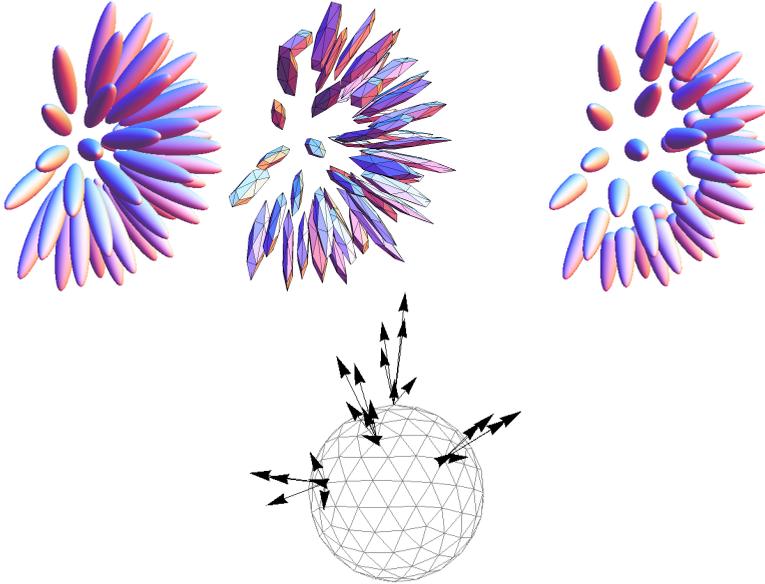


Figure 5.9: Left: Slice in \mathbb{R}^3 of the control sets (5.7) for \mathcal{F}_ε on $\mathbb{R}^3 \times S^2$, $\varepsilon = 0.2$, for different orientations of \mathbf{n} . Stencils obeying the generalized acuteness property required for Bellman type discretizations (5.46). Right: Slice in \mathbb{R}^3 of the control sets for $\mathcal{F}_\varepsilon^+$, $\varepsilon = 0.2$. Offsets used for the finite differences discretization (5.49), for four distinct orientations \mathbf{n} .

\mathbf{w}_i with its opposite. Then

$$\begin{aligned} \sum_{1 \leq i \leq 6} \rho_i (\mathbf{w}_i \cdot \mathbf{v})_+^2 &\approx (\mathbf{n} \cdot \mathbf{v})_+^2, \\ (\mathbf{n} \cdot \nabla_{\mathbb{R}^3} U(\mathbf{p}))_+^2 &\approx \frac{1}{h^2} \sum_{i=1}^6 \rho_i (U(\mathbf{x}, \mathbf{n}) - U(\mathbf{x} - h\mathbf{w}_i, \mathbf{n}))_+^2, \end{aligned} \quad (5.49)$$

up to respectively an $\mathcal{O}(\varepsilon^2) \|\mathbf{v}\|^2$ and $\mathcal{O}(\varepsilon^2 + h)$ error. Following [RT92], we design a similar upwind discretization of the angular part of the metric

$$\|\nabla_{S^2} U(\mathbf{p})\|^2 \approx (\delta_\theta U(\mathbf{p}))^2 + \frac{1}{\sin^2 \theta} (\delta_\varphi U(\mathbf{p}))^2, \quad (5.50)$$

where $\delta_\theta U(\mathbf{p})$, and likewise $\delta_\varphi U(\mathbf{p})$, is defined as

$$\delta_\theta U(\mathbf{p}) := \frac{1}{h} \max\{0, U(\mathbf{x}, \mathbf{n}) - U(\mathbf{x}, \mathbf{n}(\theta + h, \varphi)), U(\mathbf{x}, \mathbf{n}) - U(\mathbf{x}, \mathbf{n}(\theta - h, \varphi))\}.$$

We denote by $\mathbf{n}(\theta, \varphi) := (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$ the parametrization of S^2 by Euler angles $(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]$. Combining (5.49) and (5.50), one obtains an approximation of $\mathcal{F}_0^{+*}(\mathbf{p}, dU(\mathbf{p}))^2$, within $\mathcal{O}(\varepsilon^2 + r(\varepsilon)h)$ error for smooth U , denoted $\mathfrak{F}_\varepsilon U(\mathbf{p})$. We denote by $r(\varepsilon) := \max_{i=1}^6 |\mathbf{w}_i|$ the norm of the largest offset appearing in (5.48), since these clearly depend on ε . Importantly, $\mathfrak{F}_\varepsilon U(\mathbf{p})$ only depends on positive parts of

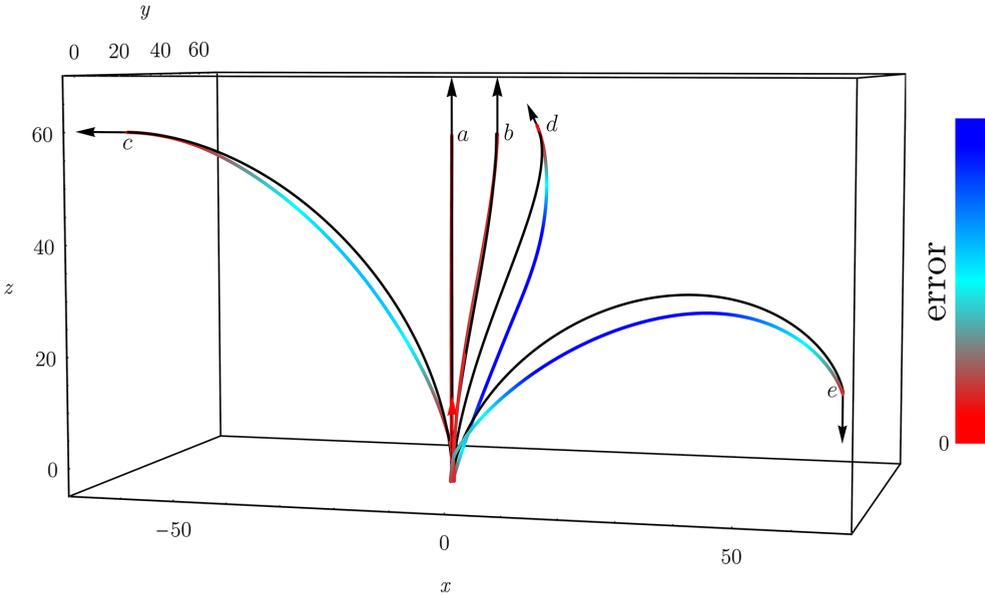


Figure 5.10: Comparison of exact geodesics (black curves) and their numerical approximation (colored curves), with $\xi = 1/64$ and $\varepsilon = .1$, for five different end conditions ($a = ((0, 0, 60), (0, 0, 1))$, $b = ((6.4, 6.4, 60), (0, 0, 1))$, $c = ((-60, 0, 60), (-1, 0, 0))$, $d = ((0, 60, 60), 1/\sqrt{6}(-1, 2, 1))$, $e = ((60, 60, 10), (0, 0, -1))$). The color indicates the error with the exact sub-Riemannian geodesics [DGDHM16].

finite differences $(U(\mathbf{p}) - U(\mathbf{q}))_+$, hence the system $\mathfrak{F}_\varepsilon U(\mathbf{p}) = 1$ can be solved using the fast-marching algorithm, as shown in [RT92]. For the convergence analysis of this discretization, as the grid scale h and tolerance ε tend to zero suitably, see [Mir17a, Mir17b].

Note that this approach could also be applied in dimension $d = 2$, and to the symmetric model $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$ featuring a reverse gear. We present only a single assessment of the numerical performance of our method, see Fig. 5.10. We compare numerically obtained shortest paths with exact SR geodesics for a small number of end points, that correspond to various types of curves. For fair end conditions (a, b, c) the numerical curves are close to the exact curves. For very challenging end conditions inducing torsion (d) or extreme curvature (e) the curves are further from the exact SR geodesics. An extensive evaluation of the performance of the numerics is left for future work.

5.7 Conclusion

We have extended the existing methodology for modeling and solving the problem of finding optimal paths for a Reeds-Shepp car to 3D and to a case without reverse gear.

Instead of using a hard constraint on the curvature as in the original paper by Reeds and Shepp [RS90], we used symmetric and asymmetric Finsler metrics. We have introduced these metrics, \mathcal{F}_0 and \mathcal{F}_0^+ , for $d = 2, 3$, such that they allow for curves that have a spatial displacement proportional to the orientation, with a positive proportionality constant in the case of \mathcal{F}_0^+ .

We have captured theoretically some of the nature of the distance maps and geodesics following from the new constrained model. We have shown in Thm. 5.2.2 that both models are globally controllable, but only the unconstrained model is also locally controllable.

The sub-Riemannian and sub-Finslerian nature is difficult to capture numerically. To this end, we introduced approximating Finsler metrics \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$, that do allow for numerical approaches. We have shown in Thm. 5.2.3 that as $\varepsilon \rightarrow 0$, the distance map converges pointwise and the geodesics converge uniformly, implying that for sufficiently small ε we indeed have a reasonable approximation of the $\varepsilon = 0$ case.

We have analyzed cusps in the metric space $(\mathbb{M}, d_{\mathcal{F}_0})$ and keypoints in the quasi-metric space $(\mathbb{M}, d_{\mathcal{F}_0^+})$ which occur on the interface surface $\partial\mathbb{M}_\pm$ given by (5.29). The analysis, for uniform costs, is summarized in Thm. 5.2.10. We have shown that cusps are absent in $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$ for $\varepsilon > 0$, that keypoints in $(\mathbb{M}, d_{\mathcal{F}_0^+})$ occur only on the boundary, and we provided analysis on how this happens. In Thm. 5.2.15 we have shown how minimizing geodesics in $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$ and $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$ can be obtained from the distance maps with an intrinsic gradient descent method.

To obtain solutions for the distance maps and optimal paths, we used a Fast-Marching method. By formulating an equivalent problem to the minimization problem for optimal paths in the form of an eikonal equation, the FM method can be applied using specific discretization schemes. We briefly compared the numerical solutions using \mathcal{F}_ε with $\varepsilon \ll 1$ with the exact sub-Riemannian geodesics in $SE(2)$ with uniform cost, which showed sufficient accuracy for not too extreme begin and end conditions.

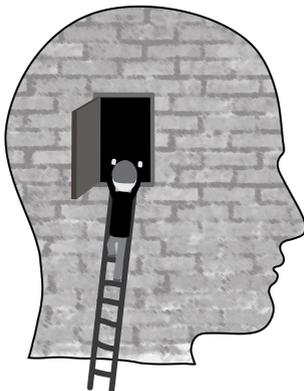
This frame work is put to practice in Chapter 7, where we show that the extension to 3D has opened up the possibility for tractography with a cost term driven by dMRI data, and that the constrained model using $\mathcal{F}_\varepsilon^+$ leads to more meaningful shortest paths in some practical cases.

Part III

Application to dMRI

Contextual PDE Flow for Improving Fiber Alignment

Based on: [PFS⁺15] J. M. Portegies, R. H. J. Fick, G. R. Sanguinetti, S. P. L. Meesters, G. Girard, and R. Duits, “Improving Fiber Alignment in HARDI by Combining Contextual PDE Flow with Constrained Spherical Deconvolution,” *PLoS ONE*, vol. 10, p. e0138122, Oct. 2015.



6.1 Introduction

Diffusion-weighted magnetic resonance imaging (dMRI) is a non-invasive technique for the characterization of biological tissue microstructure [LBBL⁺86]. Its working principles and popular processing methods have been described in Section 1.1. This chapter aims to improve tractography results for brain anatomy purposes, i.e., we want to construct an accurate estimation of white matter bundles in the brain. For the initial processing, we rely on the constrained spherical deconvolution method as described in [TCC07] to construct a fiber orientation distribution (FOD), which models the distribution of fibers along different directions. The method is able to distinguish multiple fiber directions (if present) within a single voxel, but errors and noise in the data can propagate into errors in the FOD and subsequent tractography results.

We introduce two new strategies, based on the same underlying principle, to improve fiber alignment in tractography results in order to have more reliable information on the structural connectivity of the brain. First we perform contextual regularization to the FOD obtained with CSD, see Fig. 6.1A, and secondly we introduce a fiber to bundle coherence (FBC) measure that can be applied to any fiber bundle to classify and remove spurious fibers, see Fig. 6.1B. Both approaches are based on the partial differential equation (PDE) framework discussed in Section 1.2 and Chapters 3 and 4, originally introduced in [Fra08,DF11,CDDH11,DDHCG12]. Both methods for contextual enhancement have recently been implemented by Meesters et al. in the Diffusion Imaging in Python (Dipy) software [MSG⁺16a,MSG⁺16b,GBA⁺14].

Recently, many authors [MSS13,PRD⁺10,PAV⁺15,RS12,RS13,TDV⁺14,DDHCG12,DF11,DHDT14] demonstrated the advantages of contextual processing of dMRI data. The general rationale behind contextual processing is to include alignment of local orientations and their surroundings (i.e. the context) on the coupled space of positions and orientations. For this alignment of local orientations, roto-translations are needed, which imposes a non-Euclidean structure in the PDE-based processing as we explain in Section 6.2.2. More details on the embedding of $\mathbb{R}^3 \times S^2$ in the roto-translation group $SE(3)$ can be found in Section 2.1.1 or in e.g. [DF11]. This demonstrates how either the completion or enhancement PDEs can be used to extrapolate DTI information to increase the angular resolution and resolve some fiber crossings. This idea was shown to be promising in clinical experiments [PRD⁺10,PAV⁺15], but in some cases extreme parameters had to be set to obtain clear maxima at crossings (where DTI data is inadequate). Therefore in this chapter we introduce and test the combination of CSD with contextual enhancements. The method proposed in [MSS13] uses a convection-diffusion equation to improve HARDI data to obtain connectivity measures. In our work we rely on a purely diffusive process, contour enhancement, which in contrast to contour completion

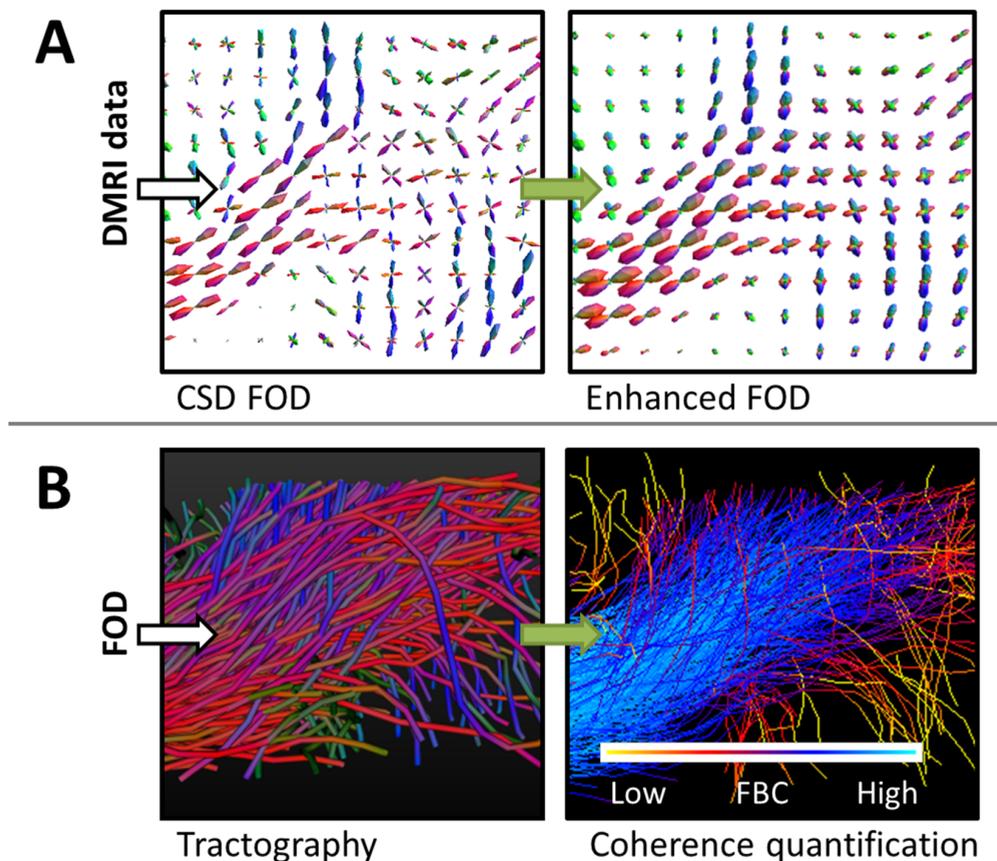


Figure 6.1: **The proposed pipeline in this chapter.** CSD is used to estimate an FOD from dMRI data. It is visualized using the glyph field visualization explained in Section 1.1. The FOD is enhanced (**A**) with PDE techniques. Then a deterministic or probabilistic tractography is applied to the (enhanced) FOD (probabilistic shown here, with coloring indicating the fiber direction). In the lower right figure, we applied our coherence quantification method (**B**), based on the same PDE framework, which shows that blue fibers are well aligned (high Fiber to Bundle Coherence (FBC)) and yellow fibers are spurious (low FBC). The green arrows indicate the steps in which the contextual PDEs are used.

does not suffer from singularities [DDHCG12] and is less sensitive to small perturbations of the initial conditions. This property makes the enhancement process more suited to be combined with the sharp angular distributions produced by CSD. As the methods mentioned above still result in broad angular distributions, they need to be combined with some sharpening method. To this end, a geometric morphological sharpening based on erosions was presented in [DDHCG12, DHDT14, TDV⁺14]. Another related method presented in [RS12, RS13] is the so-called fiber continuity model in which purely spatial

regularization is considered in combination with spherical deconvolution as alternative to the non-negativity constraint in the classical CSD [TYC⁺08]. In Section 6.2.2 we demonstrate the importance of including also an angular regularization term.

6.1.1 Contributions and outline

The first contribution of this chapter is to study the combination of the widely used CSD method with a regularization induced by the enhancement PDE acting on the FOD. Since the FOD obtained with CSD consists of sharp angular profiles, it is well-suited as an initial condition for the enhancement PDE, that typically has a smoothing effect on the orientation distributions. The contextual regularization method reduces non-aligned crossings in the FOD, allowing for a better alignment of fibers when tracking is applied on the enhanced FOD. We show that this method is therefore useful to reduce the number of false positive fibers, but mainly to find more true positives in the tractography output. Although in this chapter we compare to the classical CSD method, the PDE enhancements can also be applied to extensions of this method [TJV⁺14,SG13,JTD⁺14b,RJP⁺14,RJP⁺15].

The second contribution of this chapter is to introduce the fiber to bundle coherence (FBC) measure. The motivation for this measure is that, especially probabilistic, tracking methods typically produce spurious fibers that should be removed from the tractography. In contrast to the first approach, this method serves as a post-processing tool. For the computation of the FBC we regard the fiber bundle as a set of oriented points, by considering for every fiber point also the local tangent to the fiber. We construct a density using the enhancement PDE with an initial condition that is a sum of superposed δ -distributions at every oriented point in the bundle. The construction of such a density from tracks relates to track density imaging [CTJC10] and track orientation density imaging [DEVH⁺14], though here the use of the contour enhancement kernels as computed in Chapter 4, allows to use a sparse set of fiber tracks. The FBC, a measure for spuriousness of fibers, is computed by efficient integration of this fiber-based density. Fibers that are most spurious according to the FBC can be removed from the tractography, resulting in a better aligned fiber bundle. Complementary to the first method, this FBC measure has the purpose to remove false positives in a tractography.

Section 6.2 covers theory of the individual parts of the pipeline as outlined in Fig. 6.1, consisting of CSD, PDE enhancements, tractography and coherence quantification in Sections 6.2.1-6.2.4, respectively. In Section 6.3 we provide extensive validation of the combination of CSD and PDE enhancements and the FBC, using three experiments:

1. First we use the Tractometer evaluation system [CBG⁺12,CGB⁺13] on the ISBI 2013 HARDI reconstruction challenge dataset [DCRD⁺14], a digital phantom with

known ground truth, to demonstrate how contour enhancement improves both the local FOD reconstruction and the global connectivity of fiber bundles compared to CSD, see Section 6.3.1.

2. In Section 6.3.2 we show on a human dMRI dataset, containing different crossing bundles, that CSD combined with enhancements yields an FOD that is more robust with respect to the b -value and the number of gradient directions used in the acquisition. Furthermore, we make a comparison with earlier work involving erosions and nonlinear diffusion of FODs directly applied to a DTI-model [DDHCG12, TDV⁺14], that was based on the same data. We show that with our method the glyphs are sharper at the locations where bundles cross.
3. Finally in Section 6.3.3, we show an experiment with clinical data in which we reconstruct the optic radiation (OR) to determine the position of the tip of the Meyer’s loop, that is of interest in epilepsy surgery planning [FS63, PPA⁺05, SDBS⁺08, TDV⁺14, Meel13]. Accurate estimation of this position is difficult due to the presence of spurious fibers in the reconstruction of the OR. We show that both the FOD enhancement and the FBC measure, see Fig. 6.1, and in particular the combination of the two allow for a more stable determination of the tip of the Meyer’s loop. Here ‘more stable’ means less variation with respect to stochastic realizations in the probabilistic tractography results.

Conclusions and a discussion can be found in Section 6.4.

6.2 Methods

In this chapter we assume that we have HARDI data as input, from which we derive an FOD U that models the orientation of fibers in each voxel, i.e. $U : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}^+$. For this we use CSD [TCC12], concisely described in Section 6.2.1, as it gives sharp angular profiles and is able to distinguish multiple fiber directions within a voxel.

Then we use the enhancement PDE for diffusion of the FOD U , coupling spatial and angular information. The combination of CSD and such enhancement is a powerful method to obtain an enhanced FOD in which the coherence inherent in the data is included, providing a more coherent input for the tractography. The enhancement technique is explained in Section 6.2.2.

We use the MRtrix algorithm [TCC12] for both deterministic and probabilistic tractography to estimate the structural connectivity in the brain. In the deterministic tractography, fiber tracks are obtained by integrating a directional field, given an initial position and direction. The directional field is given by the locally maximal orientations in the glyphs. In contrast to deterministic tractography, the probabilistic tractography method

of MRtrix samples the orientations from the entire FOD and does not use just the maxima. More difficult paths can be reconstructed than with deterministic tracking, but typically also many spurious fibers are produced due to the probabilistic sampling. Both the deterministic and the probabilistic method are explained in more detail in Section 6.2.3.

In Section 6.2.4 we introduce our new technique to quantify the coherence of fibers with respect to all the fibers in a bundle, based on the same PDE theory as employed for the contextual enhancement in Section 6.2.2. We explain how the kernel of the enhancement PDE is used to construct a tractography-based density, how the FBC is computed and how this measure is able to classify spurious fibers in a tractography.

6.2.1 A brief review of CSD

We assume that at each voxel position $\mathbf{y} \in \mathbb{R}^3$ the (magnitude of the) measured signal $S_{\mathbf{y}} : S^2 \rightarrow \mathbb{R}$ can be represented by a spherical convolution of the FOD $f_{\mathbf{y}} : S^2 \rightarrow \mathbb{R}$ with a response function $K : S^2 \rightarrow \mathbb{R}$, that is estimated from the data [TCGC04]. Since the spherical deconvolution to determine the FOD is ill-posed, a non-negativity constraint is included as in [TCC07, TYC⁺08]. Then, given the signal $S_{\mathbf{y}}(\mathbf{n})$ for a sample of orientations $\mathbf{n} \in S^2$, the solution of CSD is found by iteratively solving the minimization problem:

$$f_{\mathbf{y}}^{i+1}(\mathbf{n}) = \operatorname{argmin}_{g \in \mathbb{L}_2(S^2)} \underbrace{\int_{S^2} |(K *_{S^2} g)(\mathbf{n}) - S_{\mathbf{y}}(\mathbf{n})|^2 d\sigma(\mathbf{n})}_{\text{Data Driven}} + \lambda^2 \underbrace{\int_{S^2} |(\mathcal{L}_{f_{\mathbf{y}}^i}(g))(\mathbf{n})|^2 d\sigma(\mathbf{n})}_{\text{Regularization}}, \quad (6.1)$$

for $i = 1, \dots, i_{\max}$, with i_{\max} the maximum number of iterations. Here $K \in \mathbb{L}_2(S^2)$ is aligned with and symmetric around the z -axis, the convolution $*_{S^2}$ is the usual S^2 spherical convolution [DH94], $d\sigma(\mathbf{n})$ is the Jacobian of the surface measure in orientation \mathbf{n} and λ is a parameter to influence the trade-off between the data driven term and regularization term. The linear operator $\mathcal{L}_h : \mathbb{L}_2(S^2) \rightarrow \mathbb{L}_2(S^2)$ in the regularization term gives the non-negativity constraint and is defined by:

$$(\mathcal{L}_h f)(\mathbf{n}) = f(\mathbf{n})H(\tau_h - h(\mathbf{n})), \quad \text{for given } h \in \mathbb{L}_2(S^2), \quad (6.2)$$

where H is the Heaviside function and τ_h is a threshold equal to a fixed factor τ times the mean of h . The initial function $f_{\mathbf{y}}^0$ for the iteration is computed by taking only the data driven term of Eq. (6.1). The iteration stops when successive iterations yield the same result, typically after 5 to 10 iterations [TCC07]. Throughout this chapter, we call U the FOD obtained by

$$U(\mathbf{y}, \mathbf{n}) = f_{\mathbf{y}}^{i_{\max}}(\mathbf{n}). \quad (6.3)$$

In practice CSD is performed using spherical harmonics with a maximal spherical harmonic order of 8 ($l_{\max} = 8$) as discussed in [TCC13].

Improvements to the original CSD exist to modify and improve the response function, either by recursive calibration or auto-calibration [TJV⁺14, SG13], by using multiple acquisition shells [JTD⁺14b] or by including anatomical data [RJP⁺14, RJP⁺15]. The latter two methods aim to reduce the partial volume effects, where CSD is likely to produce spurious fiber orientations. These partial volume effects can occur when in a voxel multiple tissues or multiple bundles with different orientation are present. Here we use the classical CSD as it is the basic technique available in several neuroimaging packages. However, we stress that our method is not restricted to this type of CSD. In any case, our method aims to reduce non-aligned crossings in the FOD, also the ones induced by partial volume effects. Further improvement of the methodology can be expected when including recently extended and more elaborate CSD techniques [SG13, JTD⁺14b, RJP⁺15], but this is left for future work.

6.2.2 Contour enhancement (step A)

To improve alignment of neighboring glyphs of the FOD U , recall the glyph field visualization in Fig. 6.1, we apply contextual enhancements using the diffusion equation

$$\begin{cases} \partial_t W(\mathbf{y}, \mathbf{n}, t) = (D_{33}(\mathbf{n} \cdot \nabla_{\mathbf{y}})^2 + D_{44}\Delta_{S^2}) W(\mathbf{y}, \mathbf{n}, t), \\ W(\mathbf{y}, \mathbf{n}, 0) = U(\mathbf{y}, \mathbf{n}), \end{cases} \quad (6.4)$$

that we have seen before in Chapters 3 and 4.

Although better approximations have been developed in Chapter 4, the results in this chapter are obtained using the basic approximation to the exact Green's function of the contour enhancement derived in [DF11], since at the time of writing this section for [PFS⁺15], these improved approximations were not yet at hand. This basic approximation can be written as the product of Green's functions $p_t^{\mathbb{R}^2 \times S^1}$ in the following way:

$$p_t(\mathbf{y}, \mathbf{n}) = \frac{8}{\sqrt{2}} D_{33} t \sqrt{\pi t D_{44}} p_t^{\mathbb{R}^2 \times S^1}(z/2, x, \beta) \cdot p_t^{\mathbb{R}^2 \times S^1}(z/2, -y, \gamma), \quad (6.5)$$

with $\mathbf{n} = \mathbf{n}(\beta, \gamma) = \mathbf{R}_{\mathbf{e}_x, \gamma} \mathbf{R}_{\mathbf{e}_y, \beta} \mathbf{e}_z = (\sin \beta, -\cos \beta \sin \gamma, \cos \beta \cos \gamma)^T$, $\beta \in [-\pi, \pi)$, $\gamma \in (-\frac{\pi}{2}, \frac{\pi}{2})$. The $\mathbb{R}^2 \times S^1$ kernels are given by

$$p_t^{\mathbb{R}^2 \times S^1}(x, y, \theta) = \frac{1}{32\pi t^2 D_{44} D_{33}} e^{-\sqrt{\frac{EN(x, y, \theta)}{4t}}}, \quad (6.6)$$

with

$$EN(x, y, \theta) = \left(\frac{\theta^2}{D_{44}} + \frac{1}{D_{33}} \left(\frac{\theta y}{2} + \frac{\theta/2}{\tan(\theta/2)} x \right)^2 \right)^2 + \frac{1}{D_{44}D_{33}} \left(\frac{-x\theta}{2} + \frac{\theta/2}{\tan(\theta/2)} y \right)^2. \quad (6.7)$$

To avoid numerical errors, we use the estimate $\frac{\theta/2}{\tan(\theta/2)} \approx \frac{\cos(\theta/2)}{1-(\theta^2/24)}$ for $|\theta| < \frac{\pi}{10}$. This approximation is easy to use and allows for efficient implementation [RDtHRV10].

From the approximation kernel in Eq. (6.5) it can be seen that problems could occur when either $D_{33} = 0$ or $D_{44} = 0$. To this end, a necessary and sufficient condition for the existence of a smooth solution kernel for the evolution process in Eq. (6.4) is given by the Hörmander requirement [Hör67]. This condition applies to more general situations than the one here, see e.g. [DF11], but for the specific case of contour enhancement the requirement is satisfied iff $D_{33}, D_{44} > 0$. Setting $D_{44} = 0$ would result in a singular non-smooth kernel, which has numerical disadvantages. More importantly, apart from this theoretical issue the need for both spatial and angular diffusion can also be argued from a practical point of view, as is illustrated in Fig. 6.2. We use an artificial example in which a curved fiber bundle is present, shown in the left figure. When the input is diffused with $D_{44} = 0$ as in the middle of Fig. 6.2, the peaks stay distinct and point in the wrong direction. On the other hand, when $D_{44} > 0$ as in the right figure, due to the angular diffusion the peak is redirected and the glyphs lie better aligned with the fiber bundle. Hence a positive D_{44} is needed to ensure the crucial interaction between different orientations. Finally we recall the relation between Tikhonov regularization and diffusion, see e.g. [DF11, Thm 2], which allows us to connect diffusion with $D_{44} = 0$ with the fiber continuity model in [RS12, RS13]. This model does not suffer from the inconvenience of considering only spatial regularization, as they represent the FOD in a truncated spherical harmonic basis. When the enhancements are used in combination with probabilistic tractography, we first apply a standard sharpening deconvolution transform to the FOD as described in [DDKA09], to maintain the sharpness of the FOD.

6.2.3 Tractography

As the next step in the pipeline we use the MRtrix tractography algorithm [TCC12], as implemented in <http://www.brain.org.au/software/index.html#mrtrix>, version 0.2.12. It allows us to perform deterministic and probabilistic fiber tracking on spherical harmonic representations of the (enhanced) FOD. To have a fair comparison between trackings on the FOD and the enhanced FOD, we use the parameter settings as explained next.

- In the deterministic tracking of MRtrix, seed points are randomly selected from a seed region. The initial direction is sampled randomly and every next step follows

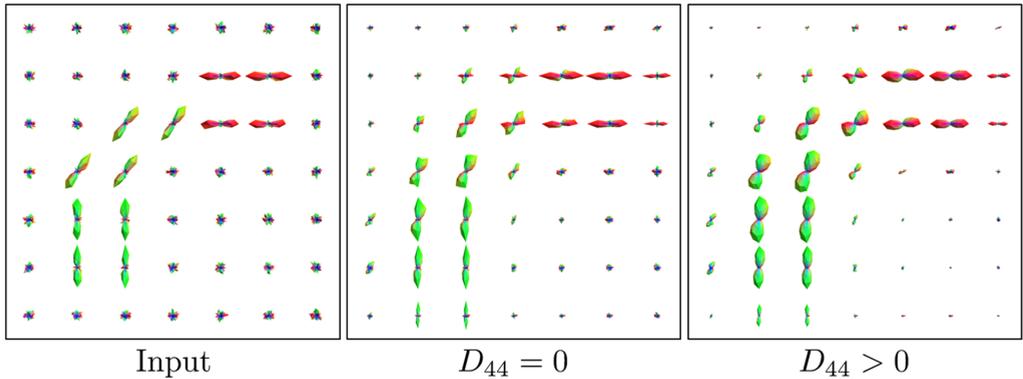


Figure 6.2: **The importance of including angular diffusion.** (Left) artificial input data of a curved bundle. (Middle) diffusion with $D_{33} > 0$, $D_{44} = 0$. (Right) contour enhancement with $D_{33}, D_{44} > 0$. Fiber propagation with $D_{44} = 0$ leads to crossing artefacts rather than smooth fiber enhancement.

the direction of the most aligned FOD maximum. If this maximum is below a threshold value, the fiber terminates. This threshold (cutoff) is set to 10% of the maximal angular response of the FOD. There is no constraint on the maximal curvature of the fibers. To prevent that fibers have an initial direction that is not aligned with the fiber bundle, we force the initial direction to be approximately in the direction of the maximal FOD peak, by setting the initial cutoff to 0.9. The step size is set to $1/10^{\text{th}}$ of the voxel size as is suggested in [TCC12]. Tracks proceed in both directions from the seed point and terminate either when they hit the boundary of the volume or mask (if applicable), or due to the threshold stopping criterion.

- In the probabilistic case, starting from the seed region, every next step follows a direction randomly sampled from the FOD. Here we set the minimal radius of curvature to 1 mm, the default value in the MRtrix algorithm. Optionally, a target region of interest is used to select only those fibers that cross this region.

The methods proposed in this chapter are not tied to the type of tractography used. In the Experiments and Results section 6.3 we validate our methods combined with both deterministic and probabilistic tractography, on synthetic and real data. In the phantom experiment described in Section 6.3.1, our preference for deterministic tractography is due to the fact that deterministic tractography was reported in [CGB⁺13] to perform better in the considered dataset than probabilistic tractography. Since only a seed region is specified and no additional information is used to filter out spurious fibers, it is difficult

to assess the quality of the probabilistic tractography results with respect to the used measures. The same holds for the experiment on real data in Section 6.3.2. On the other hand, on the optic radiation application of Section 6.3.3 where both a seed and end region are specified, a probabilistic method is required for a more complete reconstruction of the fiber bundle. Actually, in many clinical applications in which the fiber configuration can be even more complex than the phantom data we have considered, it might be preferable to use probabilistic tractography, because more information can be provided to the neurologists. Then they can decide, in combination with all of the other clinical information at their disposal, which aspects of the tractography result are to be trusted.

Streamlines from a probabilistic tractography result that are anatomically implausible can be removed with scoring methods [TDV⁺14, SDBS⁺08] or by imposing anatomical constraints. Even when using these methods, the filtered tractography output can still contain fibers that deviate from the fiber bundle and are likely to be spurious. In the next section, we propose a coherence measure for fibers in a fiber bundle in order to classify these spurious fibers.

6.2.4 Coherence quantification of fiber bundles (Step B)

In this section we introduce the second proposed processing step, a *fiber to bundle coherence* (FBC) measure to quantify the coherence of each fiber with respect to all other fibers in the bundle, recall Fig. 6.1B. A spurious fiber, as schematically shown in Fig. 6.3, is isolated from or poorly aligned with the bulk of the tracks and is therefore unlikely to represent the underlying brain structure. Fibers with low coherence, i.e. a low FBC, can then be classified as spurious.

To classify a fiber as spurious, we first construct a density by regarding each fiber as a superposition of δ -distributions in $\mathbb{R}^3 \times S^2$ and convolving this distribution with the kernel in Eq. (6.5). This density is independent of the underlying data and is based purely on the collection of fibers Γ . Integration of this density along a part of length α of a fiber gives a local measure for the coherence of that part.

Next we explain the mathematical techniques that support the idea in Fig. 6.3. We denote the fibers from a tractography output by $\mathbf{y}_i(s) \in \mathbb{R}^3$, $1 \leq i \leq N$, $0 \leq s \leq l_i$, with s the arc length parameter, l_i the total length of fiber i and N the number of fibers. Now let $\mathbf{n}_i(s) = \dot{\mathbf{y}}_i(s)$ be the tangent of the fiber, so that $\gamma_i(s) = (\mathbf{y}_i(s), \mathbf{n}_i(s))$ forms a curve (fiber) in $\mathbb{R}^3 \times S^2$. By construction, $\mathbf{n}_i(s)$ points in the forward direction of the fiber. Since in dMRI data antipodal orientations are identified, we also consider $\bar{\gamma}_i(s) = (\mathbf{y}_i(s), -\mathbf{n}_i(s))$. The complete fiber bundle is defined as $\Gamma := \{\gamma_i \mid i = 1, \dots, N\} \cup \{\bar{\gamma}_i \mid i = 1, \dots, N\}$. A discrete formulation of a fiber i with N_i points is given by:

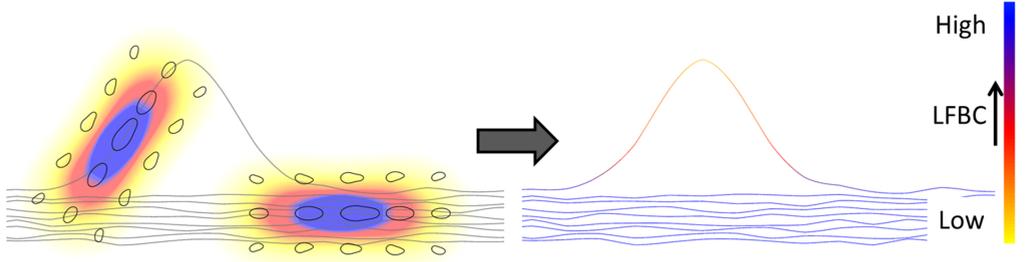


Figure 6.3: **Construction of the LFBC.** The local fiber to bundle coherence (LFBC) is constructed for a set of fibers (gray lines), illustrated in 2D for simplicity, as follows. Every local tangent in the tractography contributes to the density, by considering it as a δ -distribution. We convolve this with the contour enhancement kernel, shown on the left for two points, visualized with 2D ‘glyphs’ (polar plots), and with the coloring indicating the contribution of the kernel to the LFBC. Doing so for all points, fiber points that are isolated from or badly aligned with other fibers receive low contributions, such as the outlying fiber. The LFBC along the fibers is displayed on the right.

$$\gamma_i^j := \gamma_i(s^j) = (\mathbf{y}_i(s^j), \mathbf{n}_i(s^j)) =: (\mathbf{y}_i^j, \mathbf{n}_i^j), \quad s^j = \frac{j-1}{N_i-1} l_i, \quad j = 1, \dots, N_i. \quad (6.8)$$

This way there are $N_{tot} = 2 \sum_{i=1}^N N_i$ elements in Γ . Now we regard every point γ_i^j as a δ -distribution in $\mathbb{R}^3 \times S^2$ centered around $(\mathbf{y}_i^j, \mathbf{n}_i^j)$. A density for the entire bundle is then constructed as follows:

$$F_\Gamma(\mathbf{y}, \mathbf{n}) = \frac{1}{N_{tot}} \sum_{\sigma=1}^2 \sum_{i=1}^N \sum_{j=1}^{N_i} \delta_{(\mathbf{y}_i^j, (-1)^\sigma \mathbf{n}_i^j)}(\mathbf{y}, \mathbf{n}), \quad (6.9)$$

with index j running over points within a fiber, i running over all fibers and σ taking care of including forward and backward orientations. We use the same evolution process as in Eq. (6.4) in which $F = F_\Gamma$ now serves as initial condition, to create a diffused density $(\mathbf{y}, \mathbf{n}) \mapsto W_F(\mathbf{y}, \mathbf{n}, t)$:

$$\begin{cases} \partial_t W_F(\mathbf{y}, \mathbf{n}, t) = (D_{33}(\mathbf{n} \cdot \nabla_{\mathbf{y}})^2 + D_{44} \Delta_{S^2}) W_F(\mathbf{y}, \mathbf{n}, t), \\ W_F(\mathbf{y}, \mathbf{n}, 0) = F(\mathbf{y}, \mathbf{n}). \end{cases} \quad (6.10)$$

We solve the system in (6.10) by convolution with the corresponding kernel of (6.5), and call this the *local* FBC (LFBC):

$$\text{LFBC}(\cdot, \Gamma) = W_F(\cdot, t) = (p_t *_{\mathbb{R}^3 \times S^2} F)(\cdot), \quad (6.11)$$

with the shift-twist convolution as given in Eq. (2.23). This is illustrated in Fig. 6.3 in the 2D case. We can now define the FBC for fiber γ_i with respect to the bundle Γ as the integral of this density along the fiber:

$$\text{FBC}(\gamma_i, \Gamma) = \frac{1}{l_i} \int_0^{l_i} \text{LFBC}(\gamma_i(s), \Gamma) \, ds. \quad (6.12)$$

This results in a global property of the fiber, but spurious fibers often only locally deviate from the bundle as in Fig. 6.3. To this end, we compute for each fiber the minimum of such integrals along the fiber over intervals of length α :

$$\text{FBC}^\alpha(\gamma_i, \Gamma) = \min_{a \in [0, l_i - \alpha]} \frac{1}{\alpha} \int_a^{a+\alpha} \text{LFBC}(\gamma_i(s), \Gamma) \, ds. \quad (6.13)$$

The parameter α defines the scale over which spuriousness of fibers can be detected and is much smaller than the average fiber length. Our primary interest is not the FBC^α value itself, but rather how it compares to the average coherence of fibers in the bundle, so finally we define the *relative* fiber to bundle coherence (RFBC) as:

$$\text{RFBC}(\gamma_i, \Gamma) = \frac{\text{FBC}^\alpha(\gamma_i, \Gamma)}{\text{AFBC}(\Gamma)}. \quad (6.14)$$

Here $\text{AFBC}(\Gamma)$ is the *average* fiber to bundle coherence indicating the overall coherence of the N fibers in the bundle Γ , defined as

$$\text{AFBC}(\Gamma) = \frac{1}{N} \sum_{i=1}^N \text{FBC}(\gamma_i, \Gamma). \quad (6.15)$$

To summarize, the $\text{RFBC}(\gamma_i, \Gamma)$ of a fiber γ_i in a bundle Γ is a measure for how well aligned the least aligned part of γ_i is, compared to the average coherence of the total bundle.

In practice, we evaluate the convolution in Eq. (6.11) only in the fiber points. We compute the $\text{LFBC}(\gamma_i^k, \Gamma)$, the diffused density in the oriented point $\gamma_i^k = (\mathbf{y}_i^k, \mathbf{n}_i^k)$, recall the notation in (6.8), as follows:

$$\text{LFBC}(\gamma_i^k, \Gamma) = \frac{1}{N_{\text{tot}}} \sum_{\sigma=1}^2 \sum_{j=1}^N \sum_{q=1}^{N_j} p_t \left(\mathbf{R}_{(-1)\sigma \mathbf{n}_j^q}^T (\mathbf{y}_i^k - \mathbf{y}_j^q), \mathbf{R}_{(-1)\sigma \mathbf{n}_j^q}^T \mathbf{n}_i^k \right), \quad (6.16)$$

where $\mathbf{R}_{\mathbf{n}_j^l}$ is any rotation matrix such that $\mathbf{R}_{\mathbf{n}_j^l} \mathbf{e}_z = \mathbf{n}_j^l$, index q sums the contributions along a fiber, index j runs over all the fibers and σ as before. The FBC^α can then be computed as follows:

$$\text{FBC}^\alpha(\gamma_i, \Gamma) = \min_{a \in [0, N_i - \alpha]} \frac{1}{\alpha} \sum_{k=a+1}^{a+\alpha} \text{LFBC}(\gamma_i^k, \Gamma), \quad (6.17)$$

where $a, \alpha \in \mathbb{N}$ in this discrete case, so the LFBC is summed along short intervals of the fiber. Likewise, the AFBC can be computed as

$$\text{AFBC}(\Gamma) = \frac{1}{N} \sum_{i=1}^N \frac{1}{N_i} \sum_{k=1}^{N_i} \text{LFBC}(\gamma_i^k, \Gamma). \quad (6.18)$$

We apply this method in Section 6.3.3 for quantifying the coherence of tractography results of the optic radiation and classifying the spurious fibers.

6.3 Experiments and results

In this section we extensively test the performance of our CSD enhancement method (A) and the FBC method (B), recall Fig. 6.1 and Sections 6.2.2 and 6.2.4, in three different experiments:

- We use the HARDI Reconstruction Challenge dataset [DCDT13], which is artificial data with known ground truth, to quantitatively evaluate the CSD enhancement method (A) on deterministic tractography in Section 6.3.1.
- In Section 6.3.2 we show on dMRI human brain data that the PDE enhancement (A) has a positive effect on deterministic tractography, for different acquisition protocols of the data. Furthermore, on this dMRI dataset and on the phantom dataset we compare our method to previous work [DDHCG12], where a DTI-based FOD is used in combination with nonlinear PDE flow.
- In the third and last experiment, we reconstruct the optic radiation in human clinical data, see Section 6.3.3. We include an extensive evaluation of our methods, the enhancement of the FOD (A) and the use of the FBC to classify and remove spurious fibers (B), and the combination of both methods. We show that the reproducibility of the probabilistic tractography has increased, resulting in a more stable localization of the tip of the Meyer's loop.

For all datasets Mathematica [Wol14] was used to perform the contour enhancement algorithm and the CSD, which in practice produces the same results as the MRtrix CSD implementation when the same deconvolution kernel is used. MRtrix software [TCC12] was used to perform fiber tractography. The coherence quantification was implemented in C++. In Section 6.3.1 we make use of the Tractometer [CBG⁺12, CGB⁺13] ([http:](http://)

([//www.tractometer.org/](http://www.tractometer.org/)) to evaluate tractography results. Visualization was done in either the FiberNavigator (<https://github.com/scilus/fibernavigator>, [CWF⁺14]), Mathematica, or the open source vIST/e tool (Eindhoven University of Technology, Imaging Science & Technology Group, <http://bmia.bmt.tue.nl/software/viste/>).

6.3.1 HARDI Reconstruction Challenge

The following experiment is performed on a digital phantom dataset that was designed for the ISBI 2013 Reconstruction Challenge [DCDT13, CDD⁺14]. It is used in combination with the Tractometer [CBG⁺12, CGB⁺13], as a benchmark to compare different reconstruction and tracking methods. The phantom is inspired by the Numerical Fiber Generator [CTC⁺09] and the code to reproduce it is freely available as part of the Python package *Phantomas* (<http://www.emmanuelcaruyer.com/phantomas.php>). This synthetic dataset is of size $50 \times 50 \times 50$ voxels with a resolution of $1 \times 1 \times 1$ mm³. It consists of 27 simulated white matter bundles, designed to resemble challenging branching, kissing and crossing structures at angles between 30 and 90 degrees, with various curvature and bundle diameters ranging from 2 mm to 6 mm. An image indicating the ground truth fiber configuration is shown in the centre of Fig. 6.4.

The idea behind the signal simulation is that every voxel is subdivided into multiple sub-voxels, each one with its own attenuation profile. The final signal arrives from integrating the contribution of all the sub-voxels. Then, it is possible to combine multiple compartment types in every voxel with added Rician noise. This allows for modelling complex configurations as well as taking into account partial volume effects. While the Numerical Fiber Generator uses a tensor-like model to simulate the signal in the sub-voxels, *Phantomas* uses a CHARMED-based model [AB05]. The CHARMED model based on the Söderman-Jönsson cylinder model [SJ95] captures well the non-Gaussian behaviour of the diffusion signal for large b-values. The main reason why we selected the ISBI phantom is that it is linked with the Tractometer that allows for performing quantitative evaluations of the tractography results, using global metrics as demonstrated in the subsequent experiments.

For the experiments presented in this section we used 64 uniformly distributed gradient directions using a b-value of 3000 s/mm² with different signal to noise ratios (SNRs). We use spherical harmonics in CSD with maximal order 8, resulting in 45 estimated coefficients on each position. We then enhance the resulting FOD functions using our contour enhancement algorithm with varying parameters. From the evolutions described in Eq. (6.4) we see by a basic rescaling argument that it is sufficient to vary t and the ratio D_{33}/D_{44} . The larger this ratio, the more preference the spatial diffusion gets over the angular diffusion, resulting in elongated kernels (visualized by thin glyphs). A

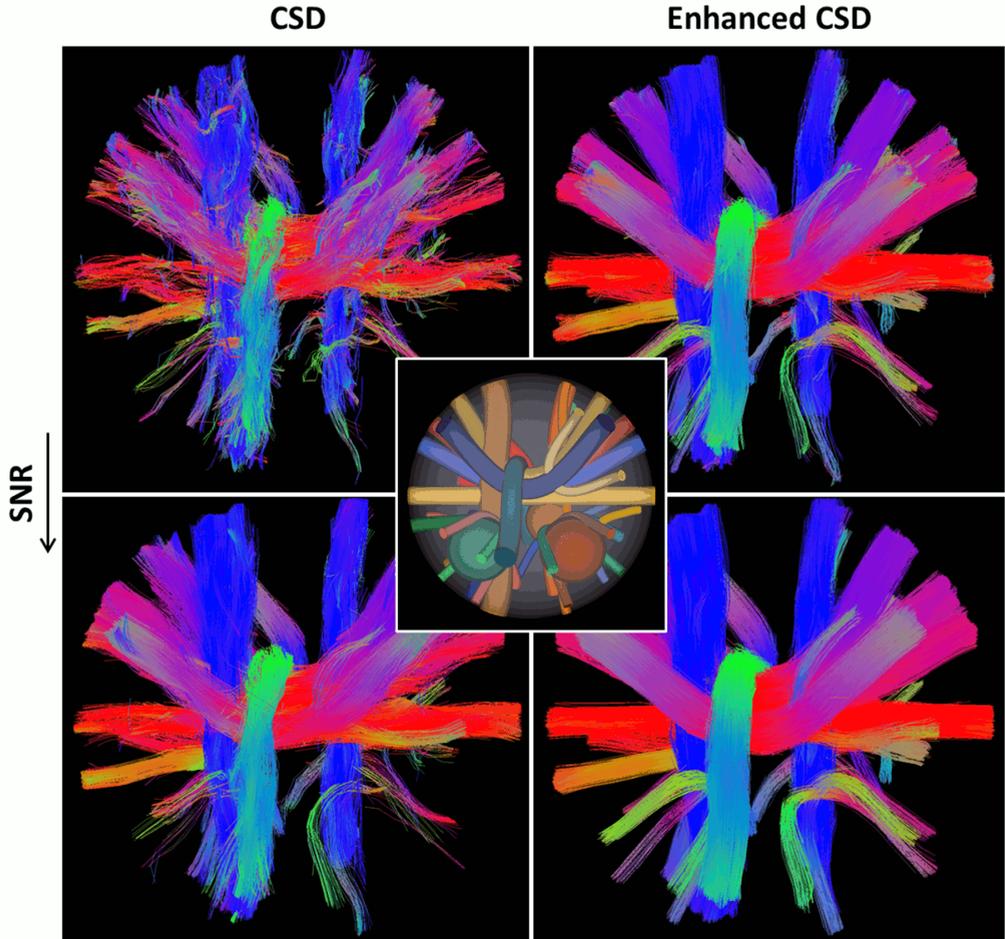


Figure 6.4: **Tractography results on the ISBI Challenge dataset.** Deterministic tractography results of CSD (left) and enhanced CSD (right) with $SNR = 2$ (top) and $SNR = 4$. The colors correspond to the direction of the fibers. The dataset consists of crossing, branching and kissing fiber bundles. The tractography on enhanced CSD results in better aligned fibers and a fuller reconstruction of the bundles. The ground truth configuration of the bundles is depicted in the center.

smaller ratio D_{33}/D_{44} is better suited in regions where the curvature of bundles is higher (visualized by thicker glyphs). The higher the diffusion time t , the more context is taken into account. When t is too large, fiber bundles with high curvature can be damaged or false positives could be created. Taking this into consideration, we choose our parameters as follows: we fix spatial diffusivity parameter $D_{33} = 1.0$, we take the angular diffusivity parameter $D_{44} \in \{0.005, 0.01, 0.02, 0.04\}$ and diffusion times $t \in [0, 5]$.

Tractography results for the entire dataset are shown in Fig. 6.4. We can recognize the positive effect of the enhancements on deterministic tractography: we see less dropouts, better aligned fibers and better continuation of fibers at crossings. An extensive quantification of the performance of our method is done at the voxel level using the FODs and at the macroscopic level using tractography in Sections 6.3.1.1 and 6.3.1.2, respectively. Both sections support the results summarized in Fig. 6.5.

6.3.1.1 Local metrics

We compare reconstructed FODs locally with the ground truth using only the orientation of the peaks. Let M be the set of voxels v in the white matter mask, then we denote the ground truth number of peaks in a voxel v by N^v and the orientations corresponding to the peaks by $\mathbf{n}_{i,\text{true}}^v$, $i = 1, \dots, N^v$.

Maxima of the constructed FOD are found by evaluating the FODs on a 60th order icosahedron tessellation with 18606 antipodally symmetric points, giving an angular resolution of less than 1 degree. Maxima are taken into account only if it exceeds a threshold of 0.1, the same value we use as threshold in the tractography. Let O_{est}^v be the set of peak orientations in voxel v estimated from the FOD. The average angular error in degrees can then be computed by:

$$\vartheta = \frac{\sum_{v \in M} \sum_{i=1}^{N^v} \min_{\mathbf{n} \in O_{\text{est}}^v} \frac{180}{\pi} \arccos(|\mathbf{n}_{i,\text{true}}^v \cdot \mathbf{n}|)}{\sum_{v \in M} N^v}. \quad (6.19)$$

In the top row of Fig. 6.5 we show the effects contour enhancement for different ratios of D_{33} and D_{44} upon variation of the diffusion time. The results are given for substantially low SNR levels 10, 6 and 4 and 2. These SNRs are computed w.r.t. the non-DW image. Specifically, if the $b=0$ intensity is 1 then the standard deviation of the Rician noise distribution is $1/\text{SNR}$. In all cases a clear improvement is found compared to CSD without enhancements and the more noise, the more the angular error is decreased. Higher diffusion times give better results and around $t = 5$ the angular error is almost stable. It can also be seen that the combination of CSD with enhancements at lower SNRs gives lower angular errors than just CSD for the higher SNRs.

There is no significant difference in the FODs between the different D_{44} values. Even though it is visible that more angular diffusion leads to fatter glyphs, for the orientation of the peaks the precise value of D_{44} is not of great importance: the angular errors for $D_{44} = 0.005$ are slightly smaller, but there is not much difference with the higher values of D_{44} .

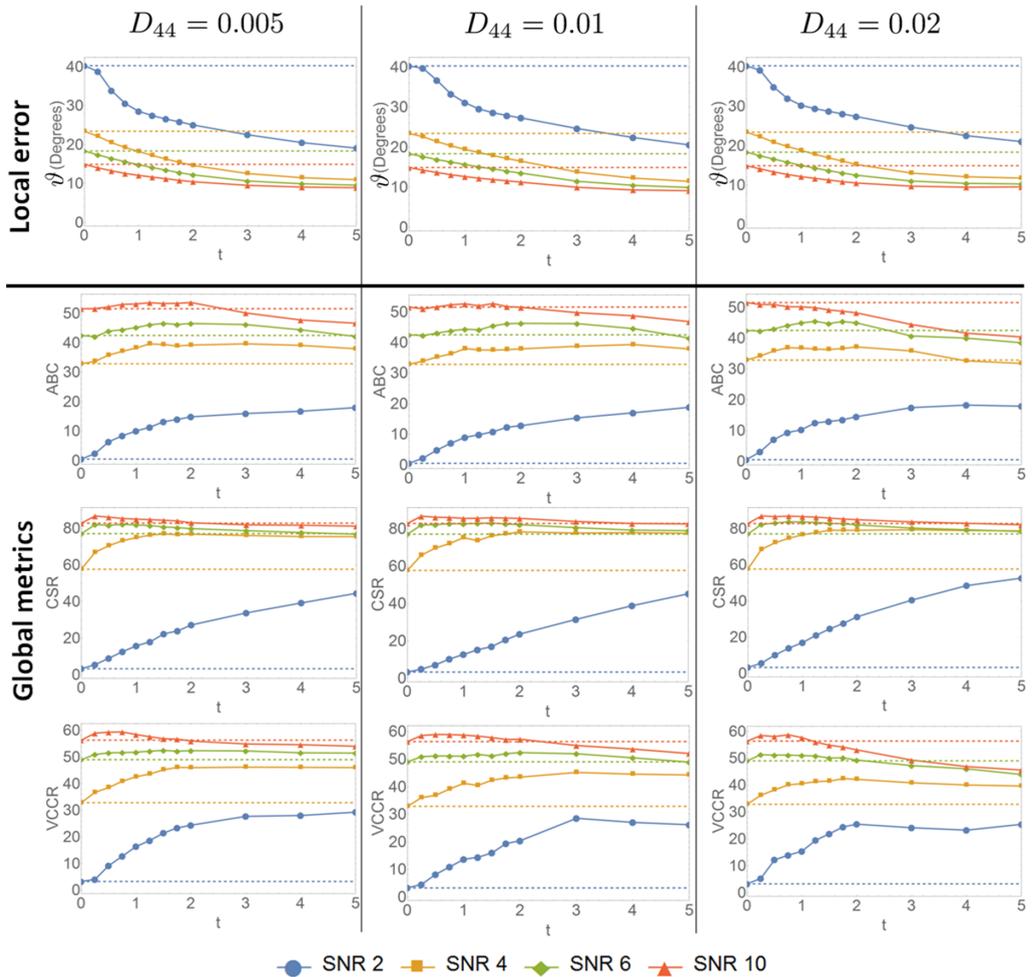


Figure 6.5: **Quantitative evaluation of the effect of enhancements.** Evolution of the local error and global metrics for three different choices of D_{44} and four different SNRs as we increase the diffusion time t . The top row shows the average angular error of the FOD peaks, the rows below show the average bundle coverage (ABC), connection to seed ratio (CSR) and the valid connection to connection ratio (VCCR), computed from the tractography results.

6.3.1.2 Global metrics

At the macroscopic level we are interested in the impact of the enhanced local reconstruction on the quality of the global connectivity. The deterministic MRtrix tractography is used as described in Section 6.2.3, with seeds randomly selected in the white matter mask. The tracks have a minimum length of 10 mm and new seed points are chosen until 10000 streamlines are selected. For every FOD, the tractography is repeated five times

with the same settings, to average out the variability in the tracking algorithm output. We then use the Tractometer [CGB⁺13] to perform a fiber tracking analysis based on the ground truth and the five results are averaged. The Tractometer outputs values for various metrics, from which we use the Valid Connections (VC), Invalid Connections (IC) and No Connections (NC). They indicate the percentage of tracks that correctly connect, incorrectly connect or do not connect gray matter areas in the dataset, respectively. We also use the Average Bundle Coverage (ABC), the percentage of voxels in a bundle that is crossed by a valid streamline, averaged over all bundles. We combine the (VC), (IC) and (NC) in two metrics introduced in [GWDD14]:

- Connection to Seed Ratio (CSR), which represents the probability that a generated fiber actually connects two gray matter areas, computed as $100\% - \text{NC}$.
- Valid Connection to Connection Ratio (VCCR), the probability that a connecting fiber is correct, computed as $\text{VC}/(\text{VC} + \text{IC})$.

The results for the ABC, CSR and VCCR with the same enhancement parameters and SNRs as for the local metric are given in Fig. 6.5. Similar remarks hold for the global metrics as for the angular error. For all three metrics and all SNRs the enhancements lead to an improvement compared to CSD, the only exception being the ABC for $\text{SNR} = 10$ and $D_{44} = 0.02$. Furthermore, as the SNR decreases, the larger diffusion times are beneficial and the more significant the improvement is. The best results are obtained for $D_{44} = 0.005$. We expect that truncation of the spherical harmonics already introduces some angular smoothing of the FODs on this artificial dataset, explaining the small effect of D_{44} in the experiments. Furthermore, we see that the diffusion time t truly acts as a regularization parameter, resulting in a robustness for the metrics with respect to the SNRs: the higher the diffusion time, the smaller the differences in the metrics between the different SNRs.

Seeding from the white matter voxels can lead to an over-representation of the number of fibers in longer fiber bundles with respect to the shorter bundles [STCC13]. The longer bundles thereby have a larger contribution to the global metrics than the shorter bundles, which could lead to an overestimation of the fiber bundles. As proposed in [STCC12], we compared the global metrics when seeding from the gray/white matter interface for CSD and one specific set of enhancement parameters. The global metrics for that seeding strategy were slightly lower for CSD and comparable when including enhancements. For the sake of comparing our enhancement method with CSD, we therefore believe it is fair to use seeding from the white matter mask.

The convincing improvement in the global metrics is supported by Fig. 6.6, that shows a selection of the fiber bundles in the dataset. It can be seen that after enhancements, there are more valid connections in the green bundle and less wrong exits in the red

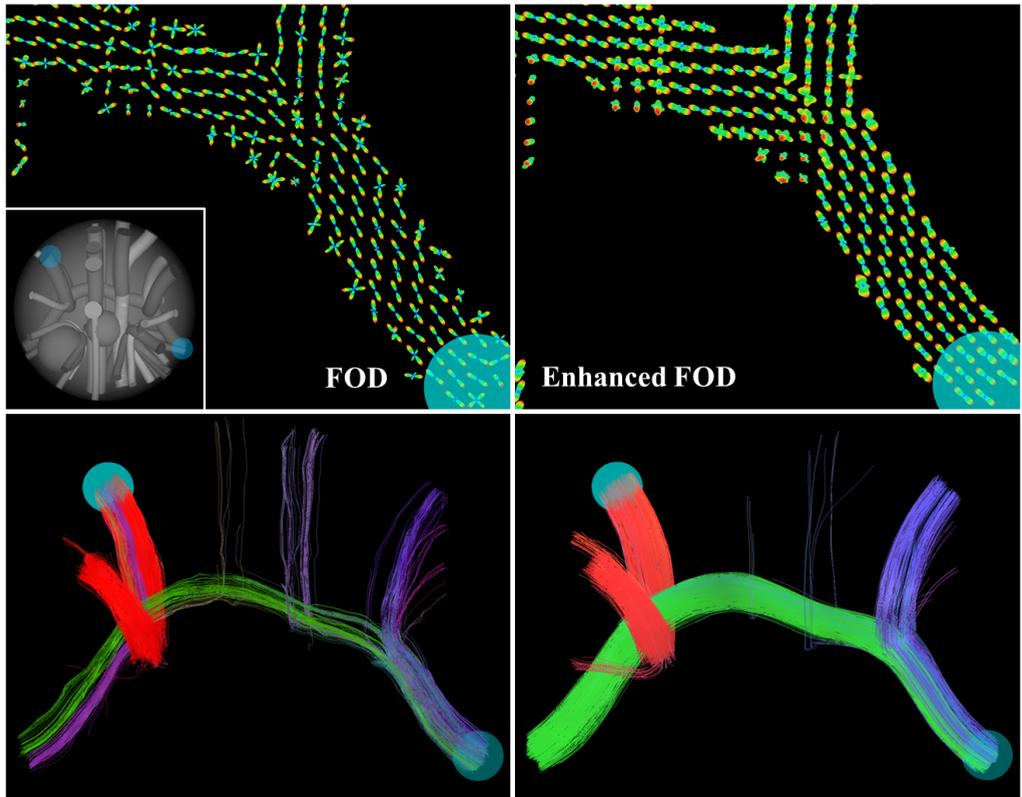


Figure 6.6: **Comparison of CSD and enhanced CSD on the ISBI dataset.** The top row shows that glyphs visualizing the FOD are better aligned, especially at the boundary of the bundle. In this case, color corresponds to the radius of the glyph. The bottom row depicts the tractography results, showing only the streamlines that pass through the indicated spheres. Here the $SNR = 4$ and the parameters used for the enhancements are $D_{33} = 1.0$, $D_{44} = 0.02$, $t = 4.0$. The ground truth image with the same viewpoint as the bottom figures is depicted on the left.

bundle, leading to a higher (VCCR) and a better bundle coverage. The glyphs in the top row show that the enhancements improve alignment of glyphs, especially at the boundary of the fiber bundles, where the original CSD result tends to suffer from partial volume effects.

6.3.2 Evaluation and comparison on dMRI data

In this experiment we consider a dMRI dataset of a part of a human brain, previously used in [DDHCG12]. The study was approved by the local ethical committee of Maastricht University, and informed written consent was obtained from the subject. Although the

dataset consists of only 10 axial slices, the corpus callosum, corona radiata and superior longitudinal fasciculus are (partly) present in the data. We show that the combination of CSD and enhancement is well-suited for different combinations of the b -value and the number of gradient directions used in the acquisition. Furthermore, we make a qualitative comparison with the DTI-based method of [DDHCG12] on this dataset and conclude with a brief quantitative comparison with this method on the dataset of 6.3.1.

6.3.2.1 Robustness with respect to the acquisition parameters

The acquisition was performed on a 3T Siemens Allegra scanner, with FOV 208x208mm and voxel size 2x2x2mm. During the data acquisition, a brain region consisting of 10 axial slices was scanned with the following combinations of b -values and N_o , the number of orientations: $b = 1000$ s/mm² with $N_o = 49$, $b = 1000$ s/mm² with $N_o = 121$ and $b = 4000$ s/mm² with $N_o = 49$. The SNR in the non-DW image was estimated to be approximately 3 using the approach of [DDL⁺11] as implemented in Dipy [GBA⁺14]. We use again CSD with spherical harmonics up to order 8. The higher b -value is obtained by using a stronger gradient pulse, making the acquisition more sensitive to detail in the tissue structure, but also inducing a lower SNR. Increasing the number of gradient directions gives a better angular resolution. We use deterministic tractography, with three seed regions manually selected in the middle of the corpus callosum, corona radiata and superior longitudinal fasciculus.

In the right column of Fig. 6.7 we show that after enhancements, the FOD allows for a more coherent reconstruction of the three bundles. Especially in the region where the three bundles come together, it can be seen that the fibers have a better propagation through the crossings. Moreover, the FODs after enhancements are very similar to each other, visible in the glyph visualization, leading to three tractography results supporting similar fiber bundles. This is an improvement with respect to CSD without enhancement, shown in the left column of Fig. 6.7. There we find more noisy FODs with more variation between the different protocols. This is also reflected in the tractography results, that contain more spurious fibers than after the enhancements.

We conclude, just like in the first experiment on the phantom data, that applying enhancements induces more robust tractography also on real dMRI data, in this case in the sense that it is less sensitive to the acquisition parameters b and N_o .

6.3.2.2 Comparison with a DTI-based FOD

In the next experiment we compare the performance of our combination of CSD with enhancements with the method in [DDHCG12] which proposed to combine DTI with *non-linear* PDE-based enhancement obtained from successively applying erosions and diffu-

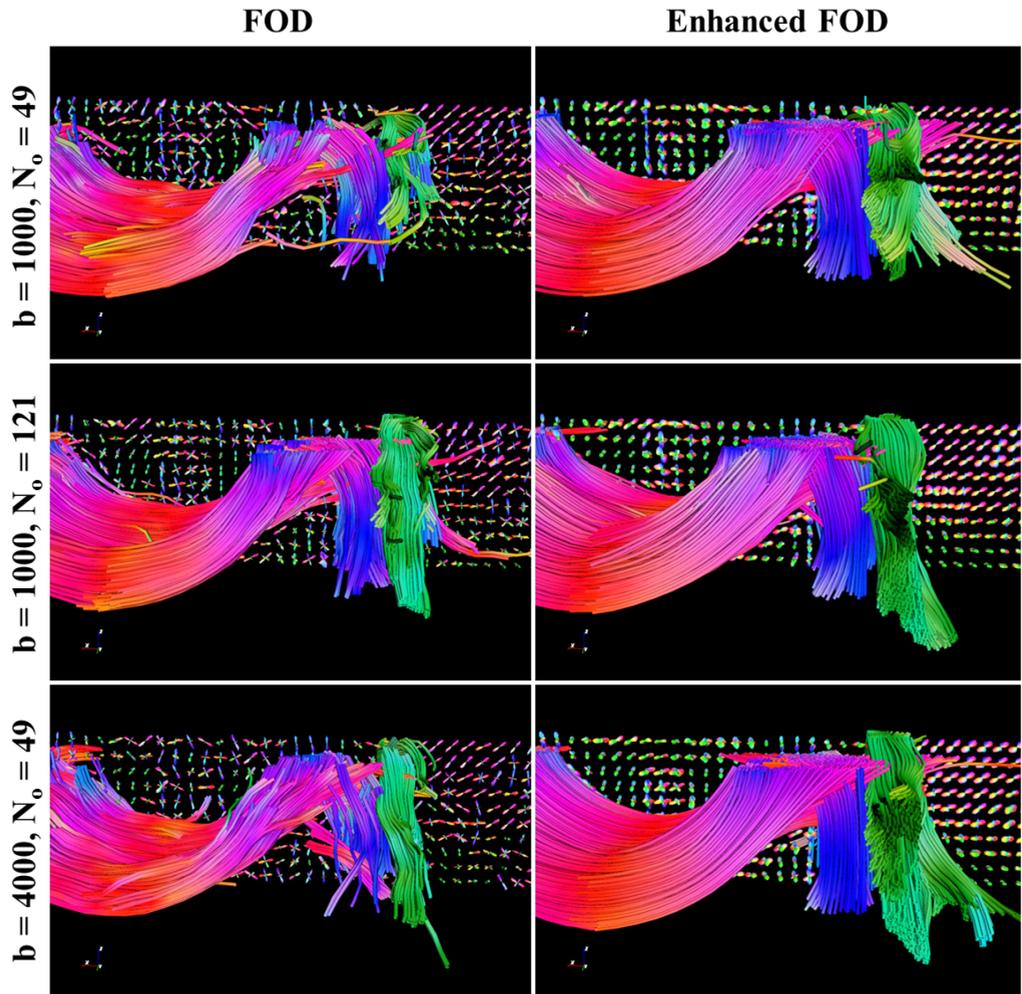


Figure 6.7: Comparison between CSD and enhanced CSD of tractography on human data. The tractography results on CSD and enhanced CSD data of the corpus callosum (mostly red), corona radiata (mostly blue) and superior longitudinal fasciculus (mostly green), with the color related to the fiber direction. Enhancements are performed with $D_{33} = 1.0$, $D_{44} = 0.02$, $t = 4.0$. All three bundles are more apparent after enhancements and more fibers pass the crossings.

sions. Let us briefly describe this method, for details we refer to [DDHCG12], and an implementation of the PDE enhancements can be found in the HARDI package for Mathematica available at (<http://bmia.bmt.tue.nl/people/RDuits/HARDIAlgorithms.zip>). First an FOD on positions and orientations that we call U_{DTI} was constructed via a transformation of the tensor field D fitted to the data [BMLB94], according to the fol-

lowing definition [ALJ⁺11, DDHCG12]:

$$U_{DTI}(\mathbf{y}, \mathbf{n}) = \frac{1}{4\pi \int_{\Omega} \sqrt{\det(D(\mathbf{x}))} d\mathbf{x}} \cdot (\mathbf{n}^T D^{-1}(\mathbf{y}) \mathbf{n})^{-\frac{3}{2}}. \quad (6.20)$$

This FOD is then sharpened with PDE erosions, a type of morphological enhancement adapted from [BBPW09], on $\mathbb{R}^3 \times S^2$ and regularized with nonlinear diffusions to find crossing structures from DTI.

Previously in [DDHCG12], the same dataset as in Fig. 6.7 for acquisition parameters $b = 1000$ s/mm² and $N_o = 49$ was processed. Here we compare the FOD obtained with CSD, that we call U_{CSD} here, with U_{DTI} in the top and bottom figures, respectively, of Fig. 6.8. Unlike DTI, which is limited by the Gaussian assumption of the diffusion profile, CSD can estimate multiple fiber orientations within a voxel. Furthermore, we see that the large glyphs in the Centrum Semiovale in the bottom figure are not apparent in U_{CSD} . Applying (linear) enhancements, as explained in Section 6.2.2, to U_{CSD} gives the second figure, and the approach in [DDHCG12] using erosions/(nonlinear) enhancements applied to U_{DTI} gives the second figure from below. It can be seen that also the enhanced DTI glyphs supports multiple fiber directions within voxels via extrapolation [PRD⁺10, DDHCG12], but at the cost of high regularization. Another noticeable difference is the fact that the glyphs in the CSD case are slimmer and crossings are more clearly defined. Whether two separate maxima are visible at a crossing is less dependent on the diffusion parameters in the PDE diffusion.

Besides the visual comparison of the FOD glyphs, we provide deterministic tractography results for both procedures in the middle of Fig. 6.8. It can be observed that both methods produce reasonable results, although the one obtained from the enhanced DTI dataset seems oversmoothed and outliers (indicated with the yellow arrow) can occur. This is due to the extreme diffusion parameters needed to perform the FOD extrapolation. We find that visually the combination of CSD and linear enhancements yields better tractography than DTI combined with erosions and nonlinear enhancements.

To provide a more quantitative and complete comparison of DTI, DTI and nonlinear enhancements, CSD and CSD with linear enhancements, we also include results of the experiment in Section 6.3.1 for the DTI methods, see Table 6.1. We heuristically determined good parameter settings for the nonlinear enhancement of DTI: erosions [DDHCG12, Eq. (59)] with $D_{11} = 0.5$, $D_{44} = 0.2$, $t = 2.0$ and diffusion [DDHCG12, Eq. (55)] with $D_{11} = 0.2$, $D_{33} = 1.0$, $D_{44} = 0.02$ and $t = 3$. In Table 6.1 is shown that applying enhancements for contextual regularization of the FOD is beneficial for both DTI and CSD. The lower the SNR, the more evident the improvements become. Furthermore, we see that in terms of the local metric, the angular error θ of the peak orientations, the DTI methods can compete with the CSD based methods. However, the global metrics

are significantly higher for CSD based methods. The quantitative results on the phantom data in Table 6.1 are in line with the qualitative comparison on real data in Fig. 6.8.

Table 6.1: For two SNR values, the results are shown for the DTI method described in Section 6.3.2, with or without nonlinear enhancements. We compare with CSD and a specific instance of enhanced CSD with parameters $D_{33} = 1$, $D_{44} = 0.01$, $t = 2$. For local metric θ lower is better, for the other metrics higher is better. In boldface are the best results for the DTI and CSD methods.

SNR 4	DTI	DTI enh	CSD	CSD enh
θ (deg.)	33.9	15.2	23.4	16.3
ABC (%)	14.3	18.1	32.9	37.9
CSR (%)	50.2	54.1	57.6	78.2
VCCR (%)	17.5	20.0	32.9	43.5
SNR 10	DTI	DTI enh	CSD	CSD enh
θ (deg.)	23.8	13.0	14.9	11.1
ABC (%)	15.5	19.9	51.6	51.5
CSR (%)	69.1	64.6	82.8	85.5
VCCR (%)	17.1	24.3	56.4	57.2

6.3.3 Improved reconstruction of the Optic Radiation

The optic radiation (OR) is a white matter fiber bundle connecting the primary visual cortex and the lateral geniculate nucleus (LGN), see Fig. 6.9. The most anterior part of the OR is called the Meyer’s loop (ML), of which the exact location is of interest for treatment of temporal lobe epilepsy [FS63, SDBS⁺08, TDV⁺14, Mee13]. During neurosurgery, a part of the temporal lobe is resected. To ensure that the OR remains intact to prevent visual field defect, it is crucial to know the distance from the tip of the Meyer’s loop to the Temporal Pole (ML-TP) [PPA⁺05], which shows large interpatient variability [NSL⁺07].

We use dMRI scans of four subjects, performed on a 3.0T Philips Achieva MR scanner, with $b = 1000$ s/mm², $N_o = 32$ and a spatial resolution of 2x2x2 mm. All subjects gave written informed consent; the study was approved by the Medical Ethics Committee of Maastricht University Medical Center (N 43386.068). The data is acquired from healthy volunteers, and ground-truth ML-TP distance is not known. Therefore accuracy of this measure of our methods cannot be checked, instead we focus on consistency and reproducibility. We apply CSD to the data to construct the FOD, with spherical harmonics up to order 6 requiring the estimation of 28 coefficients (as 32 directions are insufficient to estimate the 45 coefficients when a spherical harmonic order 8 is used, when not using super-resolution as in [TCC07]). We seed from the LGN and include all

fibers that reach the primary visual cortex. Both regions of interest are selected manually on a T1-weighted image. We use probabilistic fiber tracking as described in Section 6.2.3.

We demonstrate the effect of the enhancement of CSD and the use of the FBC measure in Sections 6.3.3.1 and 6.3.3.2, respectively, in this relevant clinical setting. A quantitative comparison of the four methods CSD (O), CSD + enhancement (A), CSD + FBC (B) and CSD + enhancement + FBC (A+B) is provided in Section 6.3.3.3. We show that the enhancement and/or the removal of spurious fibers, but in particular the combination of both methods, allows for a more stable computation of the ML-TP distance than the original tractography result.

6.3.3.1 Effect of the enhancement of CSD on tractography of the OR

In this section, we apply the PDE enhancement (step A) to the CSD FOD as before, with parameter settings $D_{33} = 1$, $D_{44} = 0.01$ and $t = 2$. After the enhancement we apply the sharpening deconvolution transform [DDKA09] and probabilistic tractography with 10000 streamlines. We compare the results of the tractography on the subjects both before and after the enhancement in Fig. 6.10. We see that the tracking on enhanced data generally shows less spurious fibers, and has a better pronounced tip of the Meyer’s loop. However, the optic radiation is a highly curved structure, where the advantage of the enhancement of elongated structures cannot be fully exploited. To further reduce the spurious fibers, we explore our other approach in the next section.

6.3.3.2 Effect of the FBC measure on tractography of the OR

In this section, we apply probabilistic tractography on subject 1, with 20000 streamlines and including state of the art data scoring as in [TDV⁺14] (only relying on the data term, i.e. $\lambda = 0$ in [TDV⁺14, Eq.(11)]), see Fig. 6.11.

The kernel parameters for the coherence quantification (step B) are set to $D_{33} = 1$, $D_{44} = 0.04$ and $t = 1.4$ for the convolution [DDHCG12]. Let Γ be the set of the 1000 most anterior fibers in a tractography of the OR, that roughly form the Meyer’s loop. We compute the LFBC and subsequently the RFBC for all the fibers in Γ .

Then we take $\epsilon_{max}(\Gamma) := \max_{\gamma \in \Gamma} \text{RFBC}(\gamma, \Gamma)$, the RFBC corresponding to the “central” fiber, in the sense that it is most coherent with the fiber bundle. We define the filtered set Γ_ϵ as

$$\Gamma_\epsilon := \{\gamma \in \Gamma \mid \text{RFBC}(\gamma, \Gamma) \geq \epsilon\}, \quad 0 \leq \epsilon \leq \epsilon_{max}. \quad (6.21)$$

This means the parameter ϵ acts as a threshold parameter and can be set such that fibers with a high spuriousness are removed. The fiber point in Γ_ϵ that is closest to the temporal pole defines the ML-TP distance. We repeat the probabilistic tractography

five times with the same settings on the same data, to qualitatively compare different stochastic realizations of the tractography method. The original OR reconstructions are shown in the top row of Fig. 6.11. We observe that due to the presence of spurious fibers, the tip of the Meyer’s loop (indicated by the orange spheres) is estimated at different locations. When we set the threshold $\epsilon = 0.1\epsilon_{max}$, removing in these cases between 6% and 8% of the most spurious fibers, we obtain the results as shown in the bottom row of Fig. 6.11. It can be seen that the resulting fiber bundles are very similar to each other, demonstrating less variation in the localization of the tip.

6.3.3.3 Quantitative comparisons on four subjects

To support our claims of the two previous sections, we test the effect of our methods on the stability of the ML-TP distance under different stochastic realizations. Here we perform probabilistic tractography with 10000 fibers ten times with the same settings, for each of the four subjects and each of the four methods (CSD, CSD + enh, CSD + FBC and CSD + enh + FBC). The FBC measure is computed from the 1000 most anterior fibers as in the previous experiment and the threshold is set to $\epsilon = 0.05\epsilon_{max}$. We compare the mean ML-TP distance and sample standard deviation determined from the tracking results of each of the methods. The results are summarized in the boxplots in Fig. 6.12. The figure strongly supports the application of the enhancements methods. For subjects 1-3 the ML-TP distance shows much less variation when including the FBC. For all subjects also (CSD + enh) gives more stable results than just CSD. Moreover, in all cases the combination (CSD + enh + FBC) outperforms CSD and for all but subject 1 the combined method (CSD + enh + FBC) also gives better results than the enhancement or FBC individually. It should be remarked that higher up the graph indicates a larger resection if used for pre-surgical evaluation, which is not necessarily positive. However, we prefer to have a stable and reproducible method that can be used with a safety margin, then a method that is more conservative, but shows large variations.

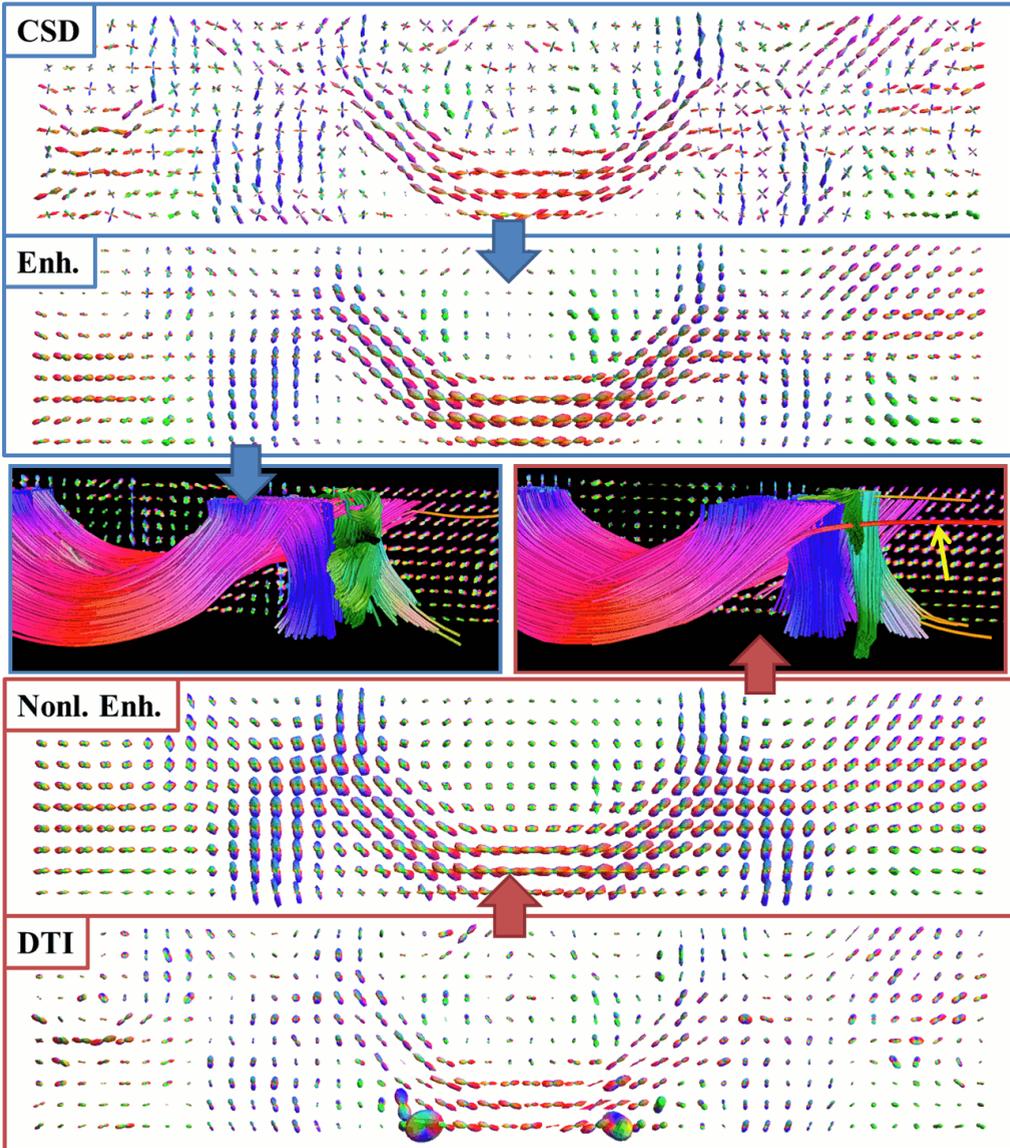


Figure 6.8: Comparison of glyph fields and tractography results between enhanced CSD and a DTI-based FOD. Glyph visualization of an axial slice of a dataset supporting the presence of the corpus callosum (mostly red), corona radiata (mostly blue) and superior longitudinal fasciculus (mostly green). Contour enhancement for CSD is performed with $D_{33} = 1.0, D_{44} = 0.02, t = 4$. Erosions and nonlinear diffusions for the DTI-based method are done with parameters as in [DDHCG12]. The tractographies corresponding to the two methods are shown in the middle. Outliers such as the red fiber, indicated by the arrow, occur due to the use of high regularization coefficients.

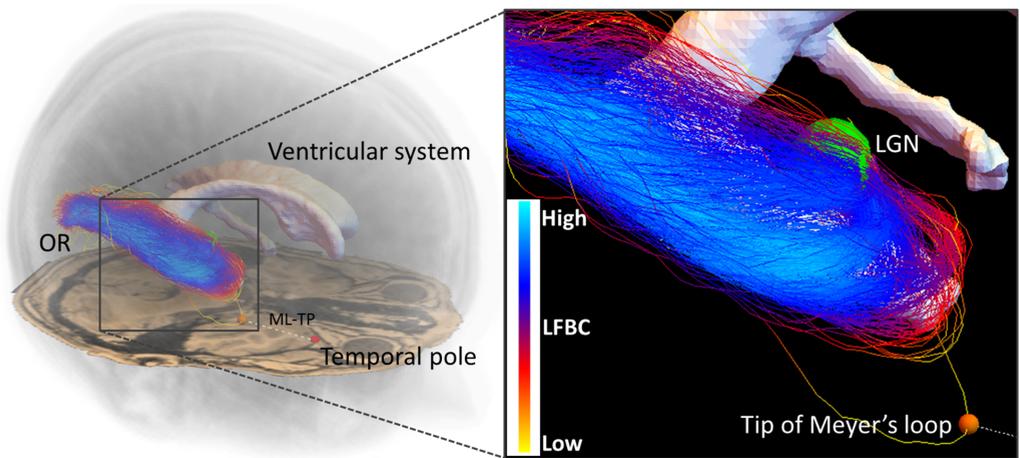


Figure 6.9: **A reconstruction of the optic radiation and its positioning in the brain.** The left figure shows how the OR is positioned in the brain, the close-up on the right shows how the OR wraps around the ventricular system. The probabilistic tractography outputs many spurious fibers. The tip of the Meyer's loop, indicated by the orange sphere, is localized on a spurious fiber and is therefore very dependent on the realization of the tractography. As a result, the distance from the Meyer's loop to the Temporal pole (ML-TP) that is used in temporal lobe resection surgery, shows a high variation among different tractography outcomes.

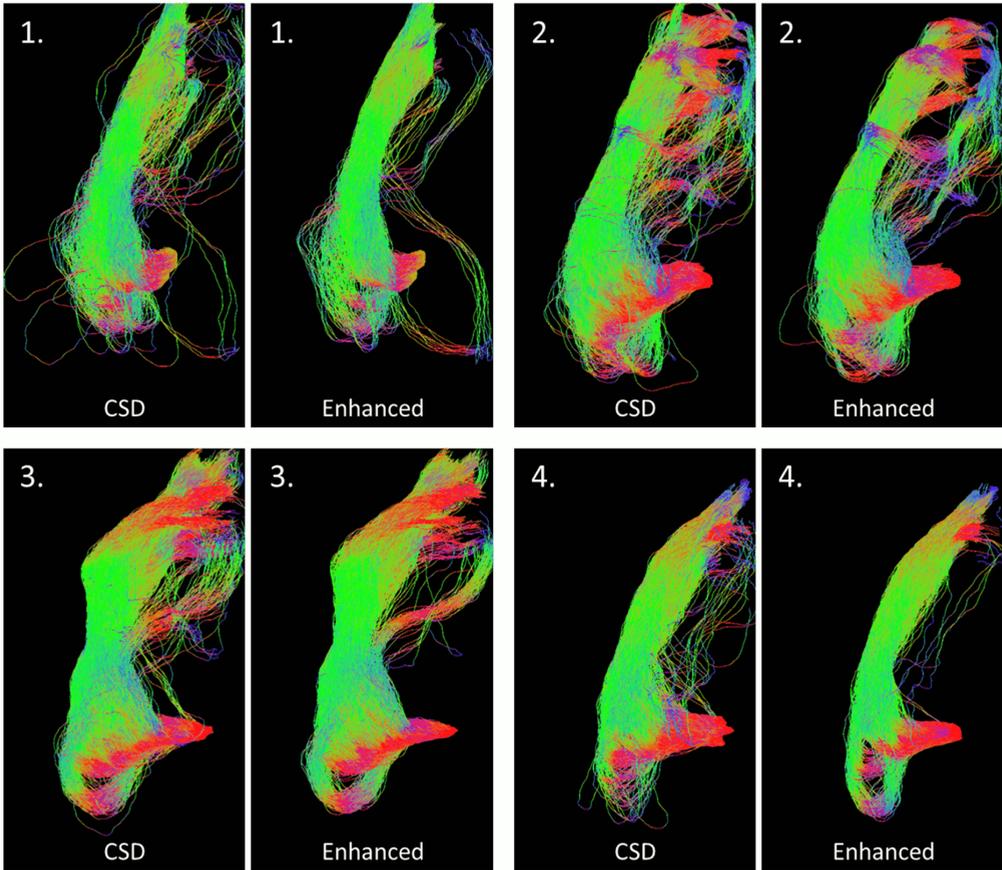


Figure 6.10: **Reconstructions of the optic radiation of four subjects with and without use of enhancements.** For all subjects, the left image shows the result on the original data, the right image shows the result on the enhanced FOD. The enhanced version generally gives less spurious fibers and has a more pronounced tip of the Meyer's loop.

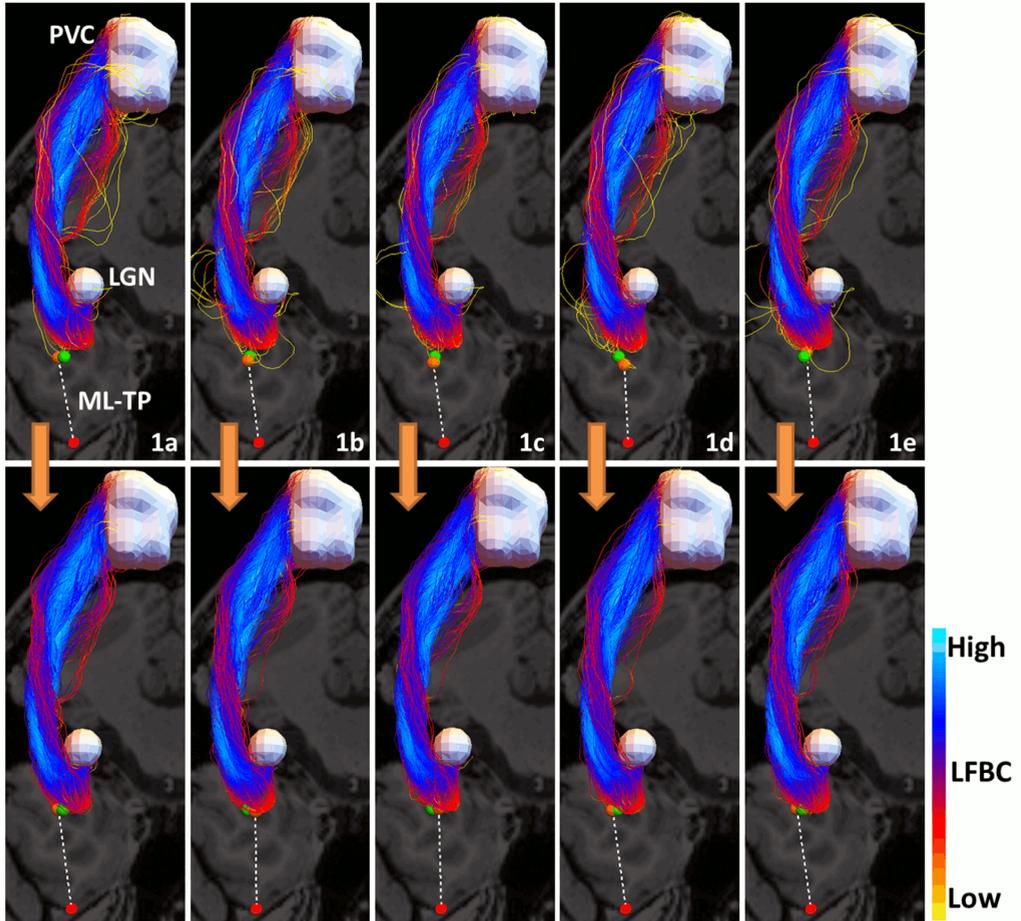


Figure 6.11: **The effect of filtering spurious fibers from a probabilistic tractography on subject 1 in five different instances.** Top row: five different instances of the probabilistic tractography of the OR, viewed from the top, selecting only the 1000 most anterior fibers. Bottom row: the result after filtering the most spurious fibers for each of the instances. The red sphere indicates the temporal pole, the white volumes represent the LGN and the primary visual cortex. The orange spheres are the positions with minimal ML-TP distance. The green sphere indicates the position of the tip averaged over the five tractography results, before (top) or after filtering (bottom). There is less variation in the position of the tip of the Meyer’s loop in the bottom row, i.e. after filtering, than in the top row. The fiber bundle in the left upper corner is the same as the one in Fig. 6.9.

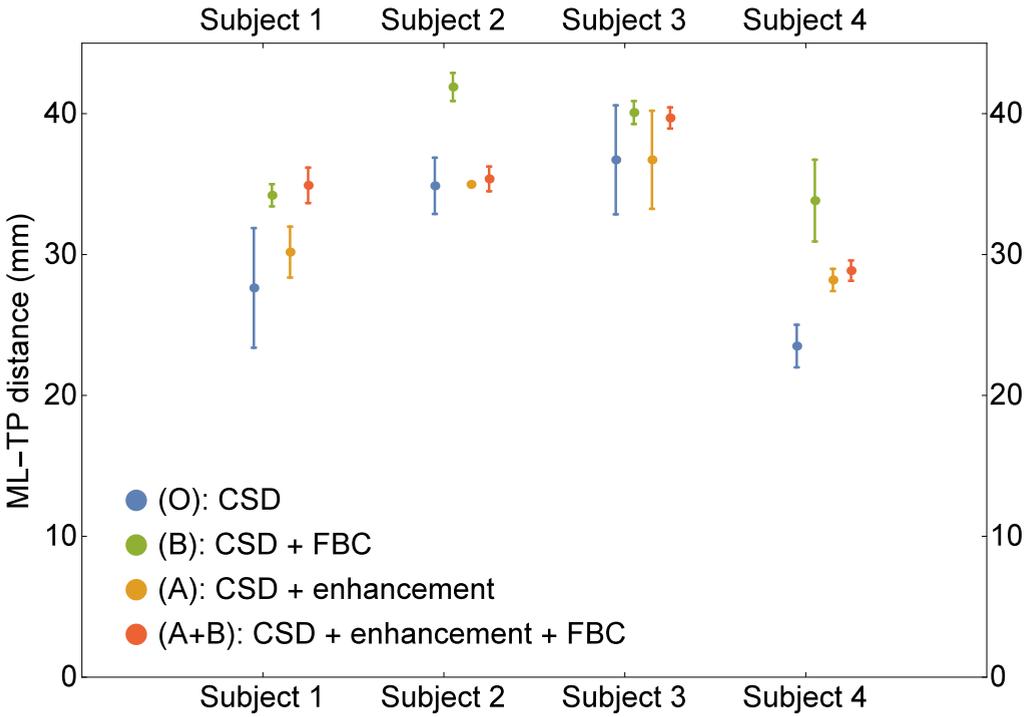


Figure 6.12: **Boxplots of the ML-TP distances.** For the four subjects, we show the mean of the ML-TP distance over ten tractography results, plus two standard deviations. The four different methods are indicated with different colors. The combination CSD + enh + FBC is the most robust in producing stable results.

6.4 Conclusion

We have proposed two new tools to improve alignment of fibers in tractography results: (A) the combination of CSD with contextual PDE enhancements and (B) a fiber to bundle coherence measure to classify spurious fibers. Both approaches rely on the same contextual processing via PDEs on the space of coupled positions and orientations. We have validated our methodology with a variety of experiments on synthetic and human data.

In the first experiment we consider a digital phantom [DCDT13] that simulates dMRI data of a challenging configuration of multiple neural-like fiber bundles for different noise levels, see Fig. 6.4. The combination of CSD with enhancements and subsequent deterministic tracking was extensively tested for varying enhancement parameters, see Fig. 6.5. The enhanced FOD peaks were compared with the ground truth fiber orientations, showing for all SNRs that the maxima of the enhanced FOD coincide better with the ground truth peaks than without application of enhancement. Also, results for data very low SNR values were much improved. To quantitatively evaluate the impact of the enhancement on the tractography results we used the Tractometer evaluation system [CGB⁺13]. The results, shown in Fig. 6.5 confirm the benefit of including the enhancement for all the metrics considered. Also an improved stability of the metrics with respect to different enhancement parameters is observed. Furthermore, we found that data with a lower SNR requires more regularization, obtained by choosing a higher diffusion time t in the enhancement. These quantitative evaluations of local and global metrics are supported by the qualitative results in Figs. 6.4 and 6.6, where we saw that after enhancement fibers are better aligned and propagate better through crossings.

The second experiment is performed on human data of a representative area of the brain with crossing fiber bundles. We evaluate our combination of CSD and enhancement for three different (single-shell) acquisition protocols, corresponding to different b -values and number of gradient directions. We observed, see Fig. 6.7, that whereas tractography on CSD without enhancement showed notable differences between the three acquisition protocols, tractography after our enhancement lead to a qualitatively similar reconstruction in all cases. This implies that the application of enhancement in the processing pipeline makes the tractography results less dependent on the scanning protocol used.

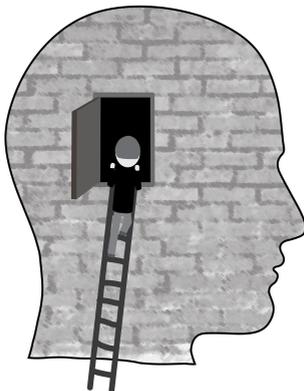
We use the same dataset and the phantom dataset to compare our method qualitatively and quantitatively with previous work [DDHCG12, TDV⁺14] in which sharpening methods and nonlinear enhancement PDEs are applied to DTI. We observed qualitatively on real data in Fig. 6.8 and quantitatively in Table 6.1 the advantage of CSD, that allows to use linear enhancements with less extreme regularization parameters than with the DTI based method, resulting in a more reliable tractography.

For our second approach to improve fiber alignment, we introduced a fiber to bundle coherence measure that can be used for detecting and filtering spurious fibers. The fiber to bundle coherence (FBC) is computed from a tractography based density that we constructed using the same PDE foundation as in the first method. As an application we considered the reconstruction of the optic radiation, a fiber bundle of which the position of the anterior extent (the Meyer's loop) is of interest for temporal lobe resection surgery. Accurate and stable localization of the tip of the Meyer's loop is difficult due to the presence of spurious fibers, as shown in Fig. 6.9. We demonstrated in Figs. 6.10, 6.11 and 6.12 that either by enhancement of the CSD FOD, or by removing the most spurious fibers using the FBC measure leads to a robust probabilistic tractography. In particular, the combination of both methods in one pipeline allows for a more stable localization of the tip of the Meyer's loop and a more stable determination of the Meyer's loop to Temporal Pole distance.

Our experiments show that our PDE enhancement methods for contextual processing are an effective and widely applicable tool to both enhance CSD data and to remove spurious fibers from tractography results. While we used CSD to construct an FOD, the PDE enhancement can be applied to an FOD obtained with any other method. We have seen that both our methods improve fiber alignment in tractography results and hence provide information on structural connectivity of the brain white matter more robustly. In the future, we aim to improve this framework by using data-adaptive smoothing, for example using local gauge frames [DJHS15].

Optimal Paths in DMRI

Partly based on: [DMMP16] R. Duits, S. Meesters, J.-M. Mirebeau, and J. Portegies, “Optimal Paths for Variants of the 2D and 3D Reeds-Shepp Car with Applications in Image Analysis,” *arXiv:1612.06137 [math]*, Dec. 2016. *Accepted for publication in JMIV, Special Issue ‘Differential Geometry and Orientation Analysis’.*



7.1 Introduction

It is generally believed that dMRI data reflects the structure of sufficiently coherent white matter fiber bundles. In particular, when representing the data in each position as a fiber orientation distribution (FOD), obtained with CSD as in the previous chapter, this FOD has (or should have) a high response in voxels in the orientation of the fibers.

The goal of tractography methods is to estimate the fiber paths, based on the (processed) dMRI data. In this chapter, we describe how we use the FOD in the shortest path optimization problem of Chapter 5, with the aim of computing shortest paths that are anatomically correct/plausible. We show on several artificial datasets, that are increasingly challenging and realistic, that our choice of metric yields promising results.

Most of the tractography methods construct tracks that locally follow the structure of the data, see e.g. [TCC12,DDKA09] or references in [JJB11]. In Chapter 6 we have shown results obtained with such local (deterministic and probabilistic) tractography methods. These methods are often bothered by local errors in the data (due to noise or artifacts), and the resulting tractogram can contain many false positives (fibers that are reconstructed but not anatomically correct) and false negatives (fibers that are present in the anatomy, but not reconstructed). Additional post-processing is often needed, with e.g. the scoring method of [TDV⁺14] or the fiber-to-bundle coherence measure of the previous chapter, to filter out an often large percentage of the fibers.

Our approach falls within the category of geodesic tracking methods, that have the advantage that they yield global minimizers of some data-dependent functional, and thereby are less sensitive to local errors in the data. These methods can be based on diffusion tensors in combination with Riemannian geometry on position space, e.g. [FJ07, LPP⁺09, JBT⁺08]. One can also make use of the more general Finsler geodesic tracking to include directionality [MMN⁺07, MPAT08], and use high angular resolution diffusion images (HARDI), examples of which can be found in [Sep11, AF09]. Recently, a promising method has been proposed, based on geodesics in the full position-orientation space using a data-adaptive Riemannian metric [PDK09]. We also work on this joint space of positions and orientations, but use either Riemannian or asymmetric Finsler metrics that are highly anisotropic.

With geodesic tractography, any two points can be connected (under mild conditions on the functional). The distance function that is computed along the way, could provide information about how well the geodesic is supported by the data and thereby also how well the two points are connected. The geodesic distance along the curve is not immediately the right measure for connectivity. Instead, we propose a connectivity measure that is a ratio between the length of the optimal path w.r.t a Finsler metric including the data, and the geodesic length of that optimal path w.r.t. a Finsler metric without

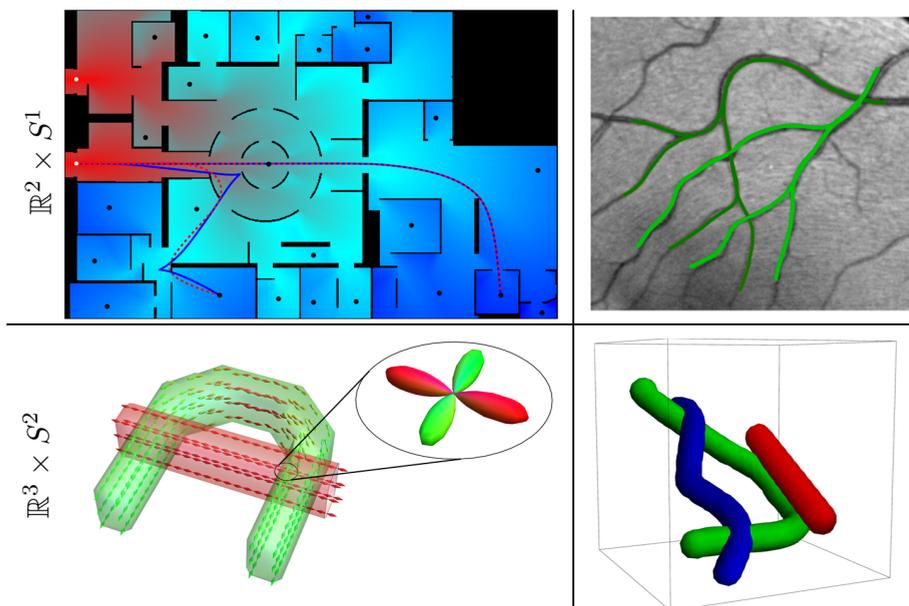


Figure 7.1: Challenges and applications. Top row: the case $d = 2$, with a toy problem for finding the shortest way with or without reverse gear (blue and red, respectively) to the exit in Centre Pompidou (top left) and a vessel tracking problem in a retinal image. Bottom row: the case $d = 3$, connectivity in (simulated) dMRI data. Left: visualization of a dataset with two crossing bundles without torsion, with a glyph visualization of the data in $\mathbb{R}^3 \times S^2$ and a magnification of one such glyph, indicating two main fiber directions. Right: the spatial configuration in \mathbb{R}^3 of bundles with torsion in an artificial dataset on $\mathbb{R}^3 \times S^2$.

the data. The full rationale behind this is discussed later.

7.1.1 The shortest path problem

In Chapter 5, we have discussed how a shortest path problem can be described as a two-step approach, first finding a distance function by solving a Hamilton-Jacobi PDE with the fast marching algorithm and then using this distance in an ODE to compute the shortest paths. The Hamilton-Jacobi approach has been shown to be useful for vessel-tracking in retinal images [BDMS15], see Fig. 7.1 (top, right). The computational advantage of the fast marching solver over the numerical method in [BDMS15] in this setting was demonstrated by Sanguinetti et al. [SBD⁺15]. A related approach using fast marching with elastica functionals can be found in [CMC16b, CMC16a]. In the approach by Bekkers et al. [BDMS15], 2D gray-scale images are first lifted to an orientation score defined on the higher dimensional manifold $\mathbb{R}^2 \times S^1$. There the combination of a sub-Riemannian metric, with the cost function derived from the orientation score, provided a

better approach to accurately track vessels in challenging sets of images. For this reason, we apply the same strategy to the field of FODs derived from dMRI data, that naturally have the domain $\Omega \subset \mathbb{R}^3 \times S^2$.

We briefly recap the essential equations for the minimization problem that we consider. For further details, see Chapter 5. On the tangent bundle of the manifold $\mathbb{M} = \mathbb{R}^d \times S^{d-1}$, $d = 2, 3$, we define a Finsler metric $\mathcal{F} : T(\mathbb{M}) \rightarrow [0, +\infty]$. We use this Finsler metric to define the length of a curve $\gamma : [0, 1] \rightarrow \mathbb{M}$:

$$\text{Length}_{\mathcal{F}}(\gamma) := \int_0^1 \mathcal{F}(\gamma(t), \dot{\gamma}(t)) dt.$$

We define the distance between two points $\mathbf{p}, \mathbf{q} \in \mathbb{M}$ as the minimal length of all connecting curves:

$$d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) := \inf \{ \text{Length}_{\mathcal{F}}(\gamma) \mid \gamma(0) = \mathbf{p}, \gamma(1) = \mathbf{q} \},$$

Our model assumption is that the curves that we consider have a certain ‘stiffness’, i.e., high curvature should be penalized. This can be achieved with either one of the following Finsler metrics (with $\mathbf{p} = (\mathbf{x}, \mathbf{n}) \in \mathbb{M}$ and $\dot{\mathbf{p}} = (\dot{\mathbf{x}}, \dot{\mathbf{n}}) \in T_{\mathbf{p}}(\mathbb{M})$):

$$\mathcal{F}_{\varepsilon}(\mathbf{p}, \dot{\mathbf{p}})^2 := \mathcal{C}_1(\mathbf{p})^2 (|\dot{\mathbf{x}} \cdot \mathbf{n}|^2 + \varepsilon^{-2} \|\dot{\mathbf{x}} \wedge \mathbf{n}\|^2) + \mathcal{C}_2(\mathbf{p})^2 \|\dot{\mathbf{n}}\|^2,$$

$$\mathcal{F}_{\varepsilon}^+(\mathbf{p}, \dot{\mathbf{p}})^2 := \mathcal{C}_1(\mathbf{p})^2 ((\dot{\mathbf{x}} \cdot \mathbf{n})_+^2 + \varepsilon^{-2} \|\dot{\mathbf{x}} \wedge \mathbf{n}\|^2 + \varepsilon^{-2} (\dot{\mathbf{x}} \cdot \mathbf{n})_-^2) + \mathcal{C}_2(\mathbf{p})^2 \|\dot{\mathbf{n}}\|^2.$$

Here $(\cdot)_+ = \max\{0, \cdot\}$ and $(\cdot)_- = \min\{0, \cdot\}$. The functions $\mathcal{C}_1, \mathcal{C}_2 : \mathbb{M} \rightarrow \mathbb{R}^+$ put a cost on spatial and angular motion, respectively, and can be chosen based on data.

These Finsler metrics model motion similar to the motion of a car: both metrics have a high penalty for moving sideways (due to the term $\varepsilon^{-2} \|\dot{\mathbf{x}} \wedge \mathbf{n}\|^2$). The difference between the two is that the Finsler metric $\mathcal{F}_{\varepsilon}^+$ also puts a high cost on moving in reverse (due to the term $\varepsilon^{-2} (\dot{\mathbf{x}} \cdot \mathbf{n})_-^2$). This avoids having cusps in the paths, those positions where the car switches from forward to backward motion or vice versa. In examples that follow later, the difference between the two models will be further clarified.

Remark 7.1.1. *The advantage of using a sub-Riemannian distance (with $\varepsilon = 0$ or practically $\varepsilon \ll 1$) over isotropic Riemannian distances has been shown in [BDMS15]. Particularly convincing results of quantitative experiments on retinal images by Erik Bekkers can be found in the supplementary materials of [BDMS15].*

An important goal in this chapter is tailoring the cost functions \mathcal{C}_1 and \mathcal{C}_2 in the Finsler metrics, such that the distance map and the optimal paths have meaning. With a number of experiments on several different datasets, we motivate the choices that we have made in that respect.

7.1.2 Overview of experiments and results

The first experiments in this chapter were carried out for the paper [DMMP16], as proof-of-concept experiments for the theory in Chapter 5. The datasets are displayed in Fig. 7.1 and consist of:

- a 2D toy example using a map of Centre Pompidou,
- a 2D retinal image,
- two synthetic Diffusion-weighted Magnetic Resonance Imaging (dMRI) datasets, with different bundle configurations.

Although Part III of this thesis was said to be about applications in dMRI, some things are just easier to grasp in 2D. We use the 2D examples, that are useful applications themselves, to point out the difference in results for the metrics \mathcal{F}_ε and $\mathcal{F}_\varepsilon^+$ ‘with and without reverse gear’, and to explain the role of the keypoints when using $\mathcal{F}_\varepsilon^+$, that occur instead of (possibly unwanted) cusps.

On the synthetic dMRI datasets we present the first application of our methods to this type of data. We present how a cost function can be extracted from the fiber orientation distributions, and how this leads to correct tracking of bundles, similar to the 2D case. The benefits of anisotropic metrics compared to isotropic metrics are demonstrated by performing geodesic tracking for various model parameter sets.

In addition to this, we show preliminary results on two dMRI benchmark datasets, see Fig. 7.2. The dataset on the left was used for the ISBI HARDI challenge 2013 [DCDT13], consisting of 27 bundles in different configurations, with data simulated using the method described in [CDD⁺14]. This dataset is used to evaluate our connectivity measure. The dataset on the right was used for the ISMRM Tractography challenge 2015 [MHNH⁺17], with dMRI data reconstructed from highly detailed and supervised tractography on Human Connectome data [VESB⁺13]. This dataset is included to show the potential on ‘almost-real’ brain dMRI data.

All computations, of both the distances and the subsequent backtracking, were performed using an anisotropic Fast Marching implementation written in C++ by Jean-Marie Mirebeau. Mathematica 11.2 (Wolfram Research, Inc., Champaign, IL) was used for further data analysis and visualization, applying Wolfram LibraryLink (Wolfram Research, Inc., Champaign, IL) to interface with the Fast Marching library.

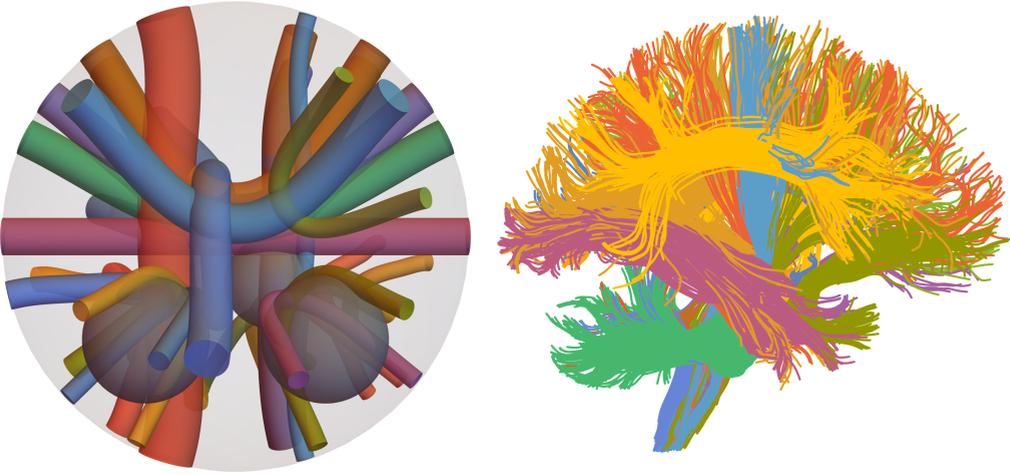


Figure 7.2: Left: ground truth bundles of ISBI HARDI challenge 2013 dataset. Right: (selection) of the ground truth fibers of the ISMRM 2015 Tractography challenge data.

7.2 Applications in 2D

7.2.1 Shortest path to the exit in Centre Pompidou

To illustrate the difference between the models with and without reverse gear and to show the role of the keypoints for non-uniform cost, we use a map of Centre Pompidou as a 2D image, see Fig. 7.3. The walls (in black) have infinite cost, everywhere else the cost is 1. We place end points (black dots) in various places of the museum and look for the shortest path from those points to one of the two exits, regardless of the end orientation. Since there are now two exits, say at \mathbf{p}_0 and \mathbf{p}_1 , the distance $U_{\mathcal{F}}(\mathbf{p})$ of any point $\mathbf{p} \in \mathbb{M}$ to one of the exits is given by

$$U_{\mathcal{F}}(\mathbf{p}) = \min\{d_{\mathcal{F}}(\mathbf{p}_0, \mathbf{p}), d_{\mathcal{F}}(\mathbf{p}_1, \mathbf{p})\}. \quad (7.1)$$

The spatial resolution is 706×441 , with 60 orientations. The cost in this example is only dependent on position, but constant in the orientation. Moreover, we use $\mathcal{C}_1 = \mathcal{C}_2$ and $\varepsilon = 0.1$.

On the left of Fig. 7.3 we see optimal paths (in blue) obtained using the Finsler metric $\mathcal{F} = \mathcal{F}_{\varepsilon}$. The fast marching algorithm successfully connects all end points to one of the exits. Some of the geodesics have cusps, indicated with white points, resulting in backward motion on (a part of) the curve. The colors show the distance $U_{\mathcal{F}_{\varepsilon}}$ as above, at each position minimized over the orientations. On the right, the optimal paths using the asymmetric Finsler metric $\mathcal{F} = \mathcal{F}_{\varepsilon}^+$ are shown in red. The curves no longer exhibit

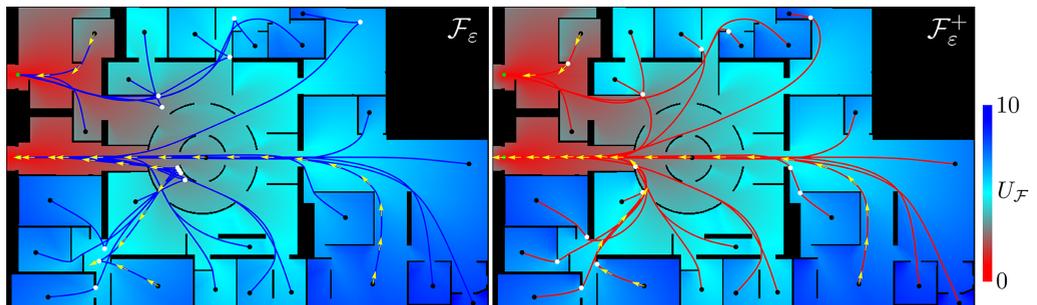


Figure 7.3: Comparison between the shortest paths from end points (black) to one of the exits (green) in a model map of Centre Pompidou, for cars with (left, blue lines) and without (right, red lines) reverse gear. The yellow arrows indicate the orientation of the curve. The background colors show the distances at each position, minimized over the orientation. White points left indicate the cusps, white points right indicate the (automatically placed) keypoints where in-place rotations take place.

cusps, but have in-place rotations (white dots) instead. These keypoints occur in this example on corners of walls. The shortest paths for this model are successions of sub-Riemannian geodesics and of in-place rotations, which can be regarded as reinitializations of the former: the orientation is adapted until an orientation is found from which the path can continue in a cusplless, optimal sub-Riemannian way.

We stress that the fast marching algorithm has no special treatment for keypoints, which are only detected in a post-processing step. We observe that keypoints are automatically positioned at positions where it makes sense to have an in-place rotation. Small differences in the distance maps between $U_{\mathcal{F}_\varepsilon}$ left and $U_{\mathcal{F}_\varepsilon^+}$ right can be observed: slightly darker colors around the corners of walls (usually in the vicinity of keypoints) indicate that the constrained model usually has a slightly higher cost there.

7.2.2 Vessel tracking in retinal images

Another application is vessel tracking in retinal images, for which the model with reverse gear and the fast marching algorithm have shown to be useful in [BDMS15, SBD⁺15]. Although the algorithm works fast and led to successful vessel segmentation in many cases, in some cases, in particular bifurcations of vessels, cusps occur. Fig. 7.4 shows one such example on the left. The image has resolution 121×114 . The cost is constructed as in [BDMS15]: the image is first lifted using cake wavelets [DFGtHR06], resulting in an image on $\mathbb{R}^2 \times S^1$. We use 64 orientations to discretize S^1 , and for the lifting and the computation of the cost function from the lifted image, we rely on standard parameter settings as in [BDMS15]. We use $\mathcal{C}_1 = \xi \mathcal{C}_2$, with $\xi = 0.02$ (top) and $\xi = 0.04$, and $\varepsilon = 0.1$. The orientations of the end conditions A, B and C (white arrows) are chosen tangent

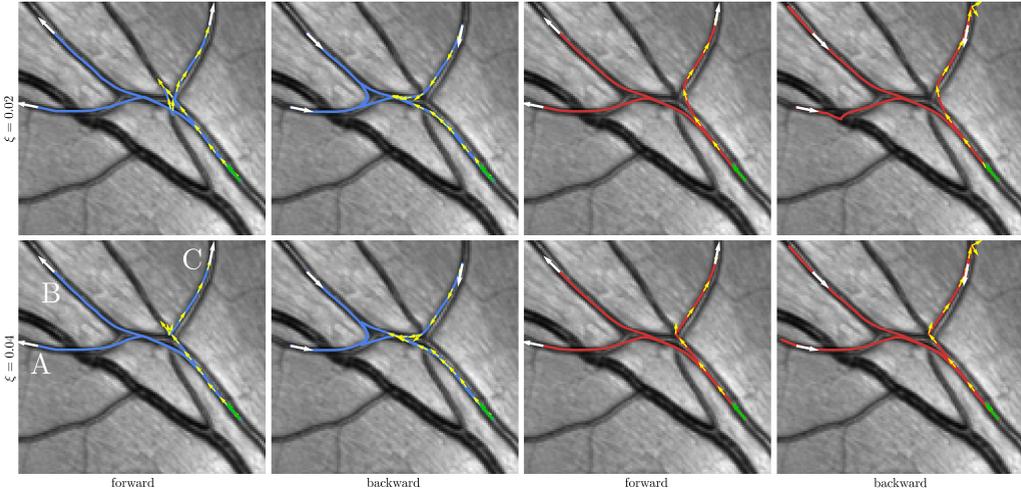


Figure 7.4: Left: shortest paths w.r.t. $d_{\mathcal{F}_\varepsilon}$, with given boundary conditions (both forward and backward). Right: shortest paths w.r.t. $d_{\mathcal{F}_\varepsilon^+}$ with the same boundary conditions. We recognize one end-condition case where on the left we get a cusp, whereas on the right we have a key-point (with in-place rotation) precisely at the bifurcation.

to the vessel, where we considered both the forward and the backward case. The vessel with end condition C is particularly challenging, since it comes across a bifurcation. For the tracking of this vessel, we indicated the orientation with yellow arrows.

The unconstrained model $(\mathbb{M}, d_{\mathcal{F}_\varepsilon})$, corresponding to the blue tracks on the left half of Fig. 7.4, gives a correct vessel tracking for the forward end conditions of A and B, for both values of ξ . This is obviously the better choice than the backward cases. However, for end condition C, neither the forward or backward with neither values of ξ yields a path without cusps. On the other hand, if we use the constrained model $(\mathbb{M}, d_{\mathcal{F}_\varepsilon^+})$, we obtain an in-place rotation or keypoint in the neighborhood of the bifurcation. Typically a higher value of ξ brings these points closer to the bifurcation. Taking the backward end conditions in combination with this model, we see in some cases that end locations are first passed by the vessel tracking algorithm, until it reaches a point where in-place rotation is cheaper, and then returns to the end position. Based on these results, we will restrict ourselves to the use of the Finsler metric $\mathcal{F}_\varepsilon^+$ in all remaining experiments of this chapter.

7.3 Application to dMRI data

7.3.1 Construction of the cost function

In our first experiments, we directly generate/simulate a Fiber Orientation Density (FOD) of a desired structure, rather than simulating actual dMRI data. We construct two simple configurations of bundles in \mathbb{R}^3 , the ones on the bottom row in Fig. 7.1. In each voxel inside a bundle, we place a spherical δ -distribution, with the peak in the orientation of the bundle. We convolve each δ -distribution with an FOD kernel that was extracted from real dMRI data and is related to the dMRI signal measured in a voxel with just a single orientation of fibers. Spherical rotation of the FOD kernel is done in the spherical harmonics domain by use of the Wigner D-matrix to prevent interpolation issues. We compose from all distributions an FOD function $U : \mathbb{M} \rightarrow \mathbb{R}^+$. This function evaluates to high values in positions/orientations that are inside and aligned with the bundle structure.

We use the FOD U to define the cost function $\mathcal{C} \geq 1$ via

$$\mathcal{C}(\mathbf{p}) = \frac{1 + \sigma}{1 + \sigma \left(\frac{U(\mathbf{p})}{\|U\|_\infty} \right)^p} \quad (7.2)$$

where $\sigma \geq 0$, $p \in \mathbb{N}$ and $\|\cdot\|_\infty$ the sup-norm. The cost function \mathcal{C} induces the following spatial and angular cost functions ($\mathcal{C}_1, \mathcal{C}_2$):

$$\mathcal{C}_1(\mathbf{p}) = \xi \mathcal{C}_2(\mathbf{p}) = \mathcal{C}(\mathbf{p})$$

The form of this non-uniform cost is comparable to the application of vessel tracking in retinal images in $d = 2$ by Bekkers et al. [BDMS15].

7.3.2 Influence of model parameters

The first synthetic dataset consists of a curved and a straight bundle, which cross at two locations as shown in Fig. 7.5. The experiments using metric $\mathcal{F}_\varepsilon^+$ demonstrate the effect of the model parameters on the geodesic back-traced from the bottom-left to the seed location at the bottom-right of the curved bundle.

Fig. 7.5 shows the resulting shortest path from \mathbf{p}_s to \mathbf{p}_e , with 4 different parameter configurations (A-D). A fixed value $p = 3$ is chosen for sufficiently sharp FODs. The shortest path in Fig. 7.5A corresponds to what we think should be the correct path. In parameter configuration B the data-term σ is lowered, which creates a geodesic that is primarily steered by internal curve-dependent costs and is shown to take the shortcut route (Fig. 7.5B). Setting $\varepsilon = 1$ in configuration C leads to a Riemannian case where the geodesic resembles a piecewise linear curve. In configuration D the relative cost of spatial movement relative to angular movement is high, leading to geodesics with shortcuts.

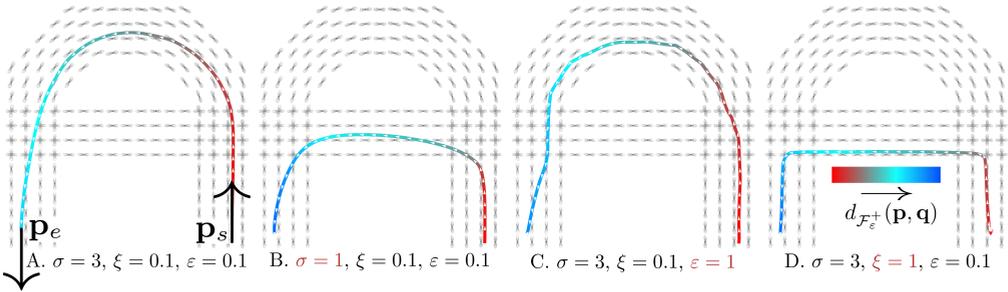


Figure 7.5: Comparison of the shortest paths from \mathbf{p}_s to \mathbf{p}_e , on a 2D plane in a synthetic dMRI dataset on $\mathbb{M} = \mathbb{R}^3 \times \mathbb{S}^2$, with four choices of parameters (A-D).

We conclude that configuration A with a relatively strong data term, large bending stiffness (ξ small), and a nearly sub-Finsler geometry ($\epsilon = 0.1$) yields the best results and avoids unwanted shortcuts.

7.3.3 Robustness to neighboring structures

Methods that provide globally minimizing curves using a dataterm often have the pitfall that dominant structures in the data attract many of the curves, much like the highway usually has the preference for cars rather than local roads. This phenomenon is to a certain extent unwanted in our applications, and we illustrate with the following example that it can be circumvented using a sub-Finsler metric on positions and orientations instead of anisotropic metric. We use the dataset as introduced in Fig. 7.1. It consists of one bundle that has torsion (green), that crosses with another bundle (blue), and a third bundle (red) that is parallel with the first in one part. The cost in these bundles is constructed in the same way as above, but now the cost in the red bundle is twice as low as in the other bundles. A small part of the data is visualized on the left of Fig. 7.6. This data is used to construct the cost function as explained above.

The spatial resolution of the data is $32 \times 32 \times 32$ and we use 162 orientations. Again we use $\mathcal{C}_1 = \xi \mathcal{C}_2 = \mathcal{C}$, with $\xi = 0.1, \sigma = p = 3$. From various positions inside the green, blue and red bundle, the shortest paths to the end of the bundles (indicated by small black arrows) computed by the FM algorithm nicely follow the shape of the actual bundles, when we choose $\epsilon = .1$ small, corresponding to an almost sub-Riemannian geodesic. This is precisely what prevents the geodesic in the green bundle to drift into the (much cheaper) red bundle. We show on the right in Fig. 7.6 that choosing $\epsilon = 1$, corresponding to having an isotropic Riemannian metric, this unwanted behavior can easily occur.

We conclude that the shortest paths w.r.t. $d_{\mathcal{F}_\epsilon}$ with $\epsilon \ll 1$, are less attracted to parallel, dominant structures than with $\epsilon \sim 1$.

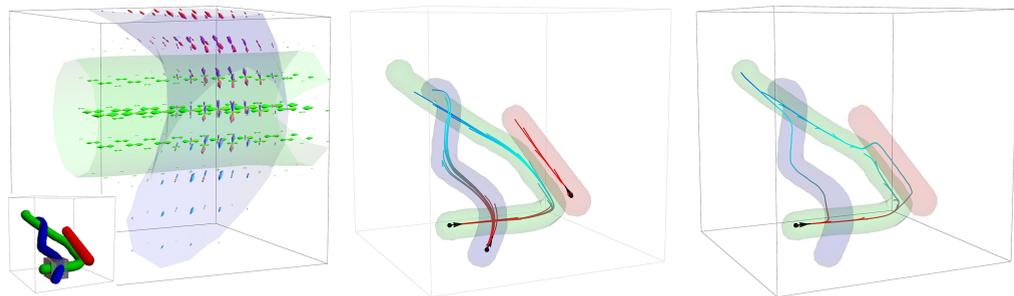


Figure 7.6: Left: 3D configuration of bundles and a visualization of part of the synthetic dMRI data. Middle: backtracking of geodesics in $(M, d_{\mathcal{F}_\varepsilon^+})$ from several points inside the curves to end points of the bundle is successful when using $\varepsilon = 0.1$. Right: when using $\varepsilon = 1$, the dominant red bundle can cause the paths from the green bundle to deviate from the correct structure.

7.3.4 ISBI HARDI Challenge 2013 dataset

Now that we have some intuition for the role of the various parameters in our Finsler metric, we move on to a more challenging dataset, the one used for the ISBI HARDI challenge 2013. This dataset was also used in Chapter 6 to show the use of contextual enhancement of the dMRI data before applying tractography. For a more detailed description of the dataset, see Section 6.3.1. In this section we want to check if we can use our shortest path problem such that we can identify the right connections.

The experimental setup is as follows. The dataset consists of 27 bundles, each with a head (H) and a tail (T), resulting in 54 points on the boundary of the sphere (where the centerlines of these bundles cross the boundary). We assume these points to be known, and we run the fast marching algorithm 54 times, each time with a different point as seed (source) point and all other points as end points. This gives us 2809 geodesics that we indicate with $\gamma_{i_X \rightarrow j_Y}$, $i, j \in \{1, \dots, 54\}$, $X, Y \in \{H, T\}$, where the cases $i = j \wedge X = Y$ are excluded.

Objective 1: optimal paths should stay inside bundle. Our first aim would be similar to the previous experiments: choose the parameters such that $\gamma_{i_H \rightarrow i_T}$ lies completely inside the ground truth volume of bundle i , to be ‘anatomically correct’. In addition to the form of the cost as in (7.2), we have made the following choices:

- Moving through an isotropic region should be expensive. To achieve this, we use on each position an \mathbb{L}_1 normalization on the sphere, i.e.

$$U_1(\mathbf{y}, \mathbf{n}) := \frac{U(\mathbf{y}, \mathbf{n})}{\int_{S^2} U(\mathbf{y}, \mathbf{n}) d\sigma(\mathbf{n})}.$$

For more easy interpretation of the data, we then divide by the global maximum:

$$U_2(\mathbf{y}, \mathbf{n}) := \frac{U_1(\mathbf{y}, \mathbf{n})}{\|U_1\|_\infty}.$$

- To increase the cost for motion through isotropic regions even further, we define an additional position-dependent (not orientation-dependent) cost \mathcal{C}_{iso} :

$$\mathcal{C}_{iso}(\mathbf{y}) = \begin{cases} 5 & \max_{\mathbf{n} \in S^2} U_2(\mathbf{y}, \mathbf{n}) \leq 0.4 \\ 1 & \text{otherwise.} \end{cases}$$

This results in additional cost for moving through regions where there are not 1 or 2 clearly distinct fiber directions according to the FOD. This is similar to a threshold on the fractional anisotropy. (Optionally, but not done here, a white matter mask could be incorporated in this position-dependent cost.)

- Finally, we use a similar shape for the cost function:

$$\mathcal{C}(\mathbf{y}, \mathbf{n}) = \mathcal{C}_{iso}(\mathbf{y}) \frac{1 + \sigma}{1 + \sigma U_2(\mathbf{y}, \mathbf{n})^p}, \quad \mathcal{C}_1 = \xi \mathcal{C}_2 = \mathcal{C}.$$

We use $\sigma = 5$, $p = 1.5$, $\xi = 0.25$.

The results of the fast marching and geodesic tracking with this choice of cost function is shown in Fig. 7.7. We conclude that the first objective can be achieved: all geodesics (in black) stay entirely inside the volume of the corresponding bundle.

Objective 2: truly connected paths should have the best connectivity We have not yet addressed the fact that with our geodesic tractography method, any two points can be connected. How should we decide which two points are actually connected? For this we introduce a connectivity measure $\kappa : (\mathbb{R}^3 \times S^2) \times (\mathbb{R}^3 \times S^2) \rightarrow [0, 1]$, that indicates for each two points, based on the connecting shortest path, how well they are connected:

$$\kappa(\mathbf{p}, \mathbf{q}) := \frac{\int_0^1 \mathcal{F}_{\mathcal{C} \equiv 1}(\gamma^*(t), \dot{\gamma}^*(t)) dt}{d_{\mathcal{F}_\mathcal{C}}(\mathbf{p}, \mathbf{q})},$$

with γ^* the shortest path with respect to the metric $\mathcal{F}_\mathcal{C}$ with \mathcal{C} as above. Using this ratio, we average out the effect of the geometry on the length of the curve, and isolate the effect of the data. Note that $\kappa = 1$ if $\mathcal{C} = 1$ along the entire geodesic γ^* , i.e., it is perfectly supported by the data, regardless of the length of the curve.

For each of the geodesics we computed the connectivity, resulting in the (symmetric) connectivity matrix displayed in Fig. 7.8, with seed points on the left (rows) and end points on the top (column). Ideally, in each row the connectivity value $\kappa(i_X, i_Y)$ should

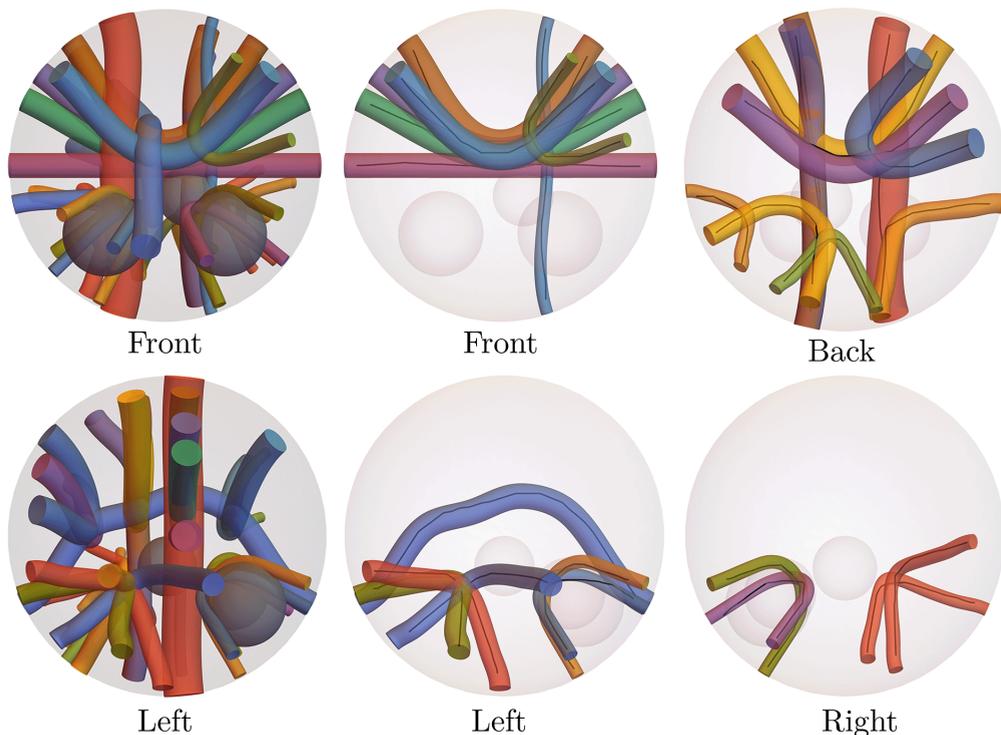


Figure 7.7: Visualization of the ISBI dataset from different viewpoints. The two left-most images show the full dataset, the other four images show a selection of the bundles. The black lines inside these bundles indicate the geodesics obtained with fast marching.

be greater than any other connectivity value in the row. If this is indeed the case, a checkmark is placed in the corresponding entry in the connectivity matrix. Otherwise, a cross is placed in the entry with the highest connectivity, and a number is placed in the entry (i_X, i_Y) that indicate the position of $\gamma_{i_X \rightarrow j_Y}$, ranked by connectivity. 21 out of 54 bundles are correctly connected. The average rank by connectivity is 3.5. However, the method performs bad on both $\gamma_{3_H \rightarrow 3_T}$ and $\gamma_{3_T \rightarrow 3_H}$. If we exclude these, the average rank by connectivity is 2.7 (1 being optimal).

When we use the assumption that a point can only be connected to one other point, we can apply a grouping algorithm, where in each iteration we take the two end points with the highest connectivity, add the corresponding geodesic to the set of reconstructed bundles, and remove the end points from the connectivity matrix. The results of this process are shown in Fig. 7.9. The green bundles correspond to correctly paired end points, the red bundles to falsely paired end points. With this grouping algorithm, 15 of the 27 bundles were correctly paired. Note that if a wrong pair is selected somewhere

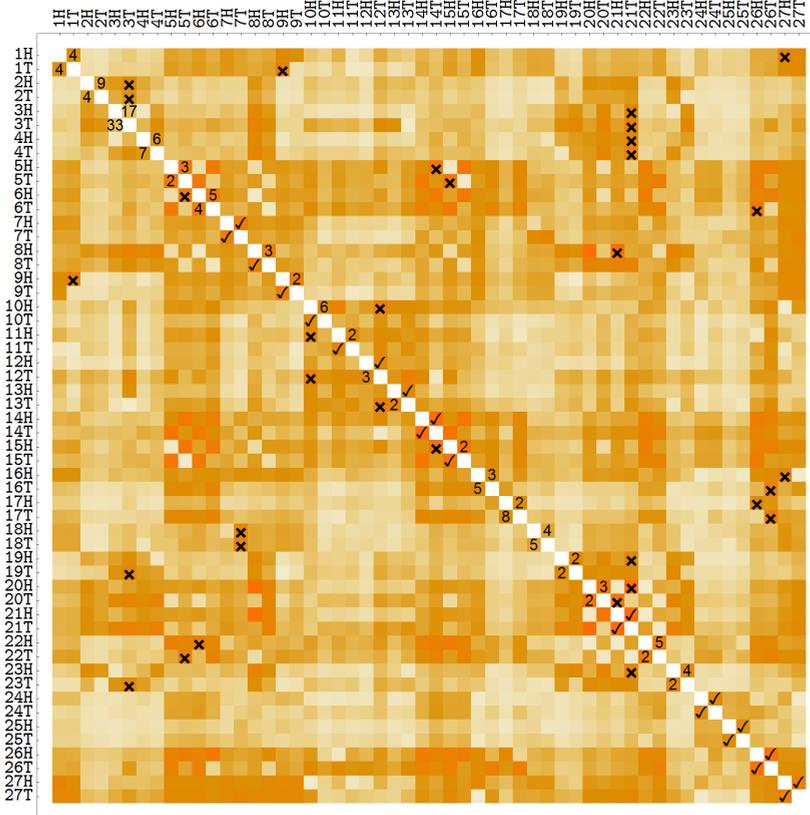


Figure 7.8: Connectivity matrix of 54 end points of the bundles in the ISBI dataset. A checkmark means the highest connectivity for the correct geodesic. A cross means the highest connectivity for an incorrect geodesic. A number means the position of that particular geodesic, ranked by connectivity.

in the course of the grouping procedure, this has negative consequences on subsequent groupings.

7.3.5 ISMRM Tractography challenge 2015 dataset

We include one last experiment on yet another dataset to prove feasibility and potential of our method for real brain dMRI data. Although the ISMRM Tractography challenge 2015 dataset is still a synthetic dataset [MNH⁺17], it is directly based on real dMRI data from the Human Connectome project [VESB⁺13].

To isolate the performance of the proposed geodesic tracking, we use the ground truth dMRI dataset instead of the noisy dataset with manually introduced artifacts that was used for the challenge itself. Testing on the corrupted data is left for future work.

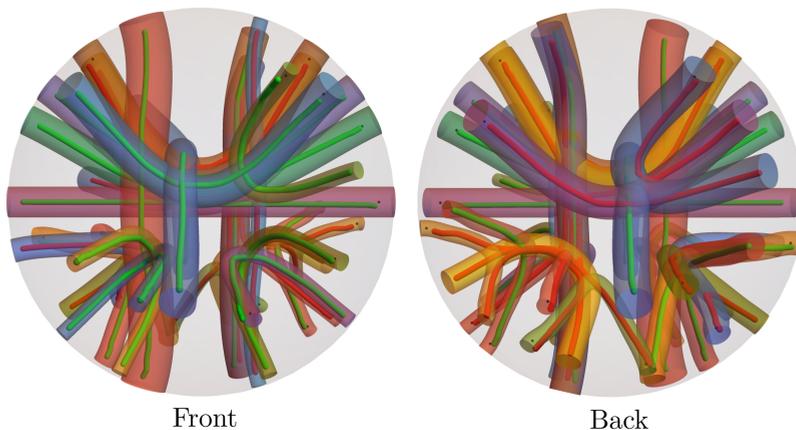


Figure 7.9: Result of pairwise grouping of endpoints by highest connectivity. The green bundles correspond to correctly grouped end points, the red bundles are incorrect.

The data has spatial dimensions $90 \times 108 \times 90$, with a voxel size of $2 \times 2 \times 2 \text{ mm}^3$ and 32 gradient directions. The dataset includes 25 white matter bundles, segmented from a global tractogram on the HCP data. We use again CSD to construct fiber orientation distributions from the diffusion data.

We limit ourselves to 4 of those 25 bundles, on the right half of the brain. The optic radiation (OR), that we also studied in Chapter 6, connects the lateral geniculate nucleus with the primary visual cortex. The other three bundles, the frontopontine tract (FPT), cortico-spinal tract (CST), parieto-occipital pontine tract (POPT) all connect the brain stem with different lobes in the cortex.

We select end points for the geodesic tracking based on the ground truth fibers. For each bundle, we apply a standard clustering of the ground truth end points, with the number of clusters as 1% of the number of fibers in that bundle. We manually place one seed point in the LGN and one in the brain stem and run the fast marching algorithm (from each seed point), with the same parameters as in the previous section.

The results of our geodesic tracking, on $\mathbb{R}^3 \times S^2$ with $d_{\mathcal{F}_e^+}$ and parameters as in the previous section, are given in Fig. 7.10. Even though we run our algorithm with just a fraction of the number of fibers in the ground truth bundle, we obtain already a reasonable reconstruction of each of the bundles, and do not observe the shortcuts (and cusps) that are often seen in geodesic tracking methods on \mathbb{R}^3 or $\mathbb{R}^3 \times S^2$. We hope to show in future experiments using the proposed connectivity measure, that such a reconstruction is sufficiently representative of the entire bundle to distinguish correct connections from incorrect ones.

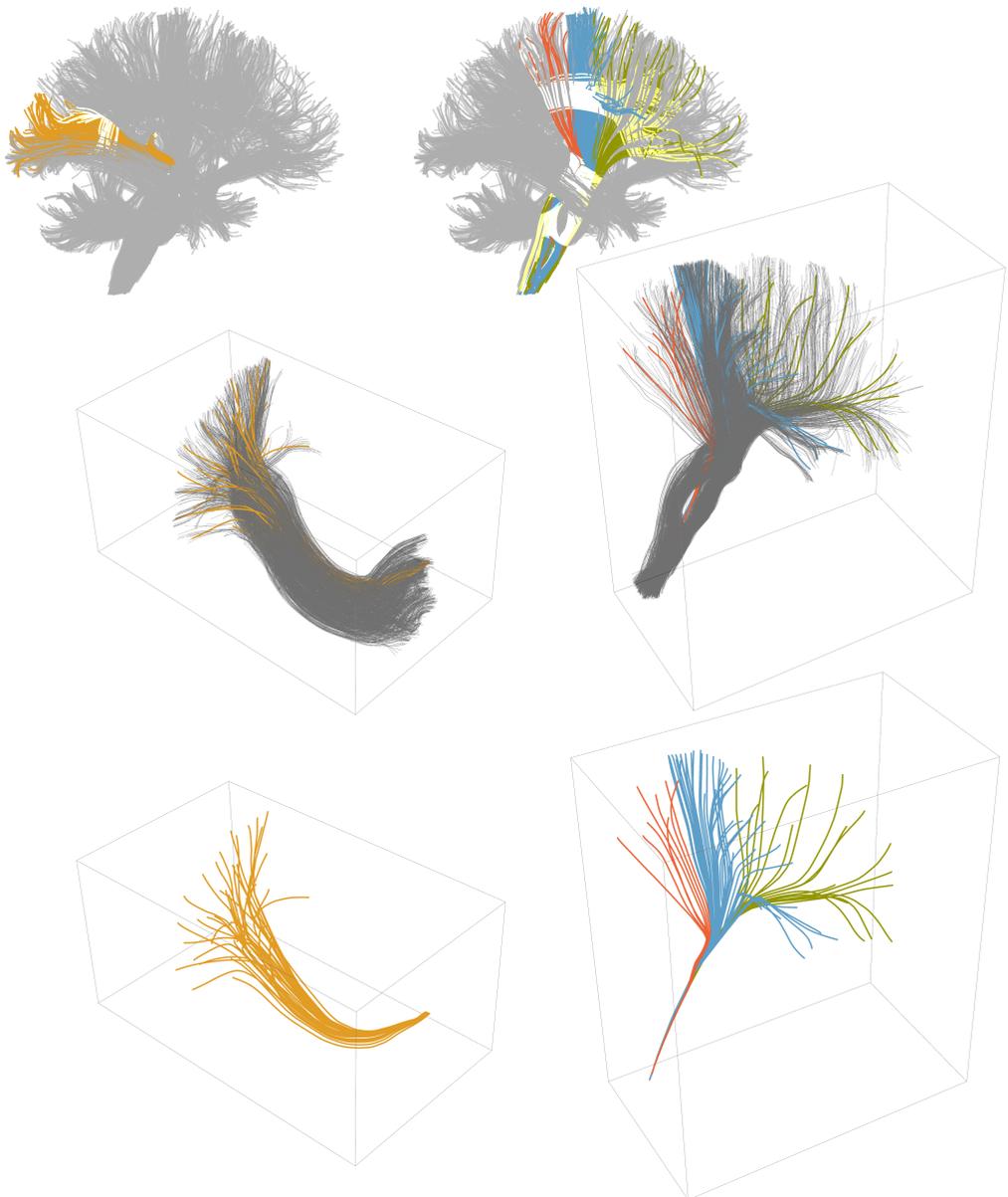


Figure 7.10: Results of our geodesic tracking for two different seed points in the right brain half. From top to bottom: location of ground truth fiber bundle(s), geodesics (color) and ground truth fibers (gray), geodesics isolated. Left: optic radiation with seed point in the right LGN. Right: frontopontine tract (green), cortico-spinal tract (blue), parieto-occipital pontine tract (red), with seed point in the brain stem.

7.4 Conclusion

In this chapter, we have proposed an approach for global geodesic tracking on dMRI data. The approach is based on the theoretical framework of Chapter 5, allowing for a fast numerical fast marching method, adapted for highly anisotropic symmetric and asymmetric Finsler metrics. The main purpose of this chapter has been to find a suitable choice for the cost functions in those metrics, to make the method data-adaptive.

Based on a number of 2D and 3D experiments, we have demonstrated the effect of the various parameters in our method. It followed from the experiment on an image of Centre Pompidou, with constant, finite cost everywhere except for the walls that, instead of having cusps when using the Finsler metric \mathcal{F}_ε , we get keypoints (in-place rotations) when using $\mathcal{F}_\varepsilon^+$. These keypoints turn out to be located on logical places in the image.

On the 2D retinal image we showed that the Finsler metric $\mathcal{F}_\varepsilon^+$ yields improved results for tackling vessel tracking through bifurcations. We see that keypoints appear close to the bifurcation, leading to paths that more correctly follow the data.

The experiments on very basic synthetic dMRI datasets show advantages of a high anisotropy and a reasonable bending stiffness (i.e., $0 < \varepsilon \ll 1$, ξ small), in the sense that the minimizing geodesics better follow the curvilinear structure and better deal with crossings and nearby parallel bundles.

Finally, we tested our approach on two benchmark dMRI datasets (ISBI reconstruction challenge [DCDT13], ISMRM tractography challenge [MHNH⁺17]). Experiments on the ISBI data indicated a reasonable performance of both the tracking and the proposed connectivity measure: by sequentially adding geodesics with the best connectivity, 15 out of 27 bundles were correctly identified, see Fig. 7.9. Preliminary experiments on the ISMRM data are very promising, see Fig. 7.10, but for now they are obtained on the noiseless dMRI data and additional research is needed on the noisy/corrupted data.

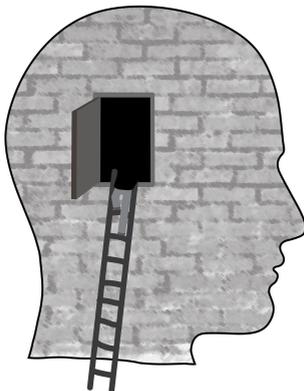
The strong performance of ‘car modeling’-Finsler metrics in 2D vessel tracking and positive first results on artificial dMRI data, encourages us to pursue a more quantitative assessment of the performance in both 3D vessel tracking problems (in for example Magnetic Resonance Angiography) and in actual dMRI data.

We will also attempt to improve the proposed connectivity measure, by defining a measure for brain regions instead of single end points. Furthermore, including a sliding window (integrating over a part of the geodesic instead of the entire geodesic), similar to the FBC measure in Chapter 6, can lead to considerable improvements. This way, paths that are inappropriate on an intermediate length scale (and therefore should be discarded) result in a worse connectivity than is currently the case.

Finally, we aim for an even more data-adaptive vessel/fiber tracking, by integrating locally adaptive frames [DJHS16] into the Finsler metrics \mathcal{F}_ε , $\mathcal{F}_\varepsilon^+$.

Holistic Image Reconstruction for Diffusion MRI

Based on: [GPG⁺16] V. Golkov, J. M. Portegies, A. Golkov, R. Duits, and D. Cremers, “Holistic Image Reconstruction for Diffusion MRI,” in *Computational Diffusion MRI*, pp. 27–39, Springer, 2016.



8.1 Introduction

Among the main problems in diffusion MRI (dMRI) are scan duration limits (thus a limited amount of data), image resolution limits, noise, and image artifacts. In recent years, a host of methods [DF10a,LDP07,KHK⁺12,PMG⁺15,TTS⁺14,FGB⁺14,CXB⁺15,MJG⁺15,RML⁺14] have been developed to tackle these issues. These methods use (simplified) assumptions about the data, such as specific types of smoothness / transform-domain sparsity / low-rankedness, specific types of data similarity between different coordinates in the 3-D space of diffusion directions and weightings (q -space), accurate or simplified image acquisition models, in some cases combined with a tailored acquisition strategy.

Super-resolution in dMRI allows increasing the resolution beyond the hardware limits. In the original super-resolution techniques for dMRI [SGW12,PJB⁺13], there is no coupling of different q -space coordinates, *i.e.* each q -space coordinate is treated independently without taking advantage of common structure. It is performed from image space to image space, independently of the image reconstruction step. Recent methods [TNRMH14,GSM⁺15,VSJV⁺15] couple q -space coordinates and use the original data-acquisition space but regularize only in the reconstruction space – not in additional spaces.

The proposed method allows leveraging complementary information by coupling in q -space, while imposing data consistency in the original space and balancing regularization in several arbitrary representations simultaneously.

The rest of this chapter is organized as follows. In Section 8.2.1, we describe the data formation model. In Section 8.2.2, we introduce holistic reconstruction (raw data consistency, several regularization spaces, super-resolution reconstruction in image and diffusion space) and give details on sampling in acquisition and reconstruction spaces, the regularizers, the optimization procedure and its implementation. We show results of holistic super-resolution reconstruction after artificial subsampling of Human Connectome Project data in Section 8.3 and conclude with a discussion in Section 8.4.

8.2 Methods

8.2.1 Image acquisition model

The image is modeled on a domain $\Omega \times \mathbb{R}^3$, where $\Omega \subset \mathbb{R}^3$ represents the domain in image space, and dimensions four to six of $\Omega \times \mathbb{R}^3$ represent the space consisting of three-dimensional diffusion directions and diffusion weightings (q -space) for which discrete

samples are acquired. A complex-valued dMRI image ρ is a mapping

$$\rho : \Omega \times \mathbb{R}^3 \rightarrow \mathbb{C} \text{ given by} \quad (8.1)$$

$$(\mathbf{y}, \mathbf{q}) \mapsto \rho(\mathbf{y}, \mathbf{q}) = r(\mathbf{y}, \mathbf{q}) \exp(i\varphi(\mathbf{y}, \mathbf{q})), \quad (8.2)$$

where r is the image magnitude and φ is the image phase at spatial coordinate $\mathbf{y} \in \Omega$ and q -space coordinate $\mathbf{q} \in \mathbb{R}^3$. Magnitude r and phase φ are mappings

$$r : \Omega \times \mathbb{R}^3 \rightarrow \mathbb{R}, \quad (8.3)$$

$$\varphi : \Omega \times \mathbb{R}^3 \rightarrow S^1. \quad (8.4)$$

These images are not acquired directly. Acquisition is performed in k -space (more precisely: in the joint six-dimensional (k, q) -space), after Fourier transform $\mathcal{F}_{1,2}$ along the spatial dimensions 1 and 2 of Ω . When sampled at N data points, the resulting data $\mathbf{d} \in \mathbb{C}^N$ forms from r and φ according to

$$\mathbf{d} = T(r, \varphi) + \varepsilon, \quad (8.5)$$

where ε is complex-valued i.i.d. Gaussian noise (thermal noise) and T is the encoding operator. The operator T composes r and φ pointwise into a complex-valued image via $C(r, \varphi) = r \odot \exp(i\varphi)$ where “ \odot ” is the pointwise product, followed by a Fourier transform into (k, q) -space and discrete sampling S :

$$T(r, \varphi) = S\mathcal{F}_{1,2}C(r, \varphi), \text{ with} \quad (8.6)$$

$$S : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{C}^N \text{ given by} \quad (8.7)$$

$$(S\hat{\rho})_n = \int_{[-0.5, 0.5]^3} \hat{\rho}(k_n + v, q_n) dv, \quad (8.8)$$

where the $((k_n, q_n))_{n \in \{1, \dots, N\}}$ are the sampling points in (k, q) -space. Details can be found in refs. [Val14, BCH⁺14].

8.2.2 Holistic reconstruction

Our goal is to reconstruct the image magnitude r and phase φ from the acquired data \mathbf{d} . In order to improve image quality, such a reconstruction should include state-of-the-art image processing methods, such as denoising, super-resolution reconstruction and orientation distribution function¹ (ODF) enhancement. Rather than performing this in a classical manner, where each step is performed separately, we couple all transformations and regularizers into a single optimization problem. This allows performing the

¹The ODF is a formalism that characterizes the strength of diffusion in different directions. It is defined formally below in Eq. (8.10).

entire reconstruction in a single step, while having full control over the balance between all regularizers simultaneously. Furthermore, this avoids data-consistency formulations in intermediate spaces, where the noise distribution is difficult to model correctly (*e.g.* Rician signal distribution and other cases) – our least squares data term penalizes deviation from k -space measurements, where noise is Gaussian, while still reconstructing and regularizing in arbitrary spaces. Finally, a holistic formulation allows regularizing in additional spaces other than the acquisition and the reconstruction space. This allows for example using information from the ODF (otherwise calculated independently at a later step) to inform the super-resolution reconstruction in image space.

In our proof-of-concept holistic reconstruction experiments, we treat the entire six-dimensional data jointly (rather than treating each q -space coordinate independently during image space reconstruction, followed by treating each image coordinate y independently during q -space-based processing) and combine the following concepts into a single optimization problem:

- Data consistency in the original (k, q) -space,
- Reconstruction into (y, q) -space with super-resolution in both the spatial and diffusional dimensions,
- Spatial regularization of (y, q) -space data,
- Angular regularization of (y, q) -space data by treating each q -space shell independently as functions on the (uncoupled) space $\mathbb{R}^3 \times S^2$ of positions and orientations,
- Spatial and angular regularization of the ODFs which implicitly correspond to the reconstructed (y, q) -space data by treating them as functions on the (uncoupled) space $\mathbb{R}^3 \times S^2$ of positions and orientations.

The general form of holistic reconstruction into (y, q) -space is

$$\arg \min_{r, \varphi} \frac{1}{2} \|T(r, \varphi) - d\|^2 + R(r), \quad (8.9)$$

where $R(r)$ is a sum of regularization terms which may or may not transform the image magnitude r into another space, such as ODFs, prior to penalizing non-regularity².

The “codomain” of our pipeline, *i.e.* the reconstruction space, can be extended into diffusion models, as in refs. [WDAH13, VBK13]. These model-based methods can be complemented by our regularizers in additional spaces to yield a holistic framework.

²The precise formula that we use for $R(r)$ will follow later in Eq. (8.12).

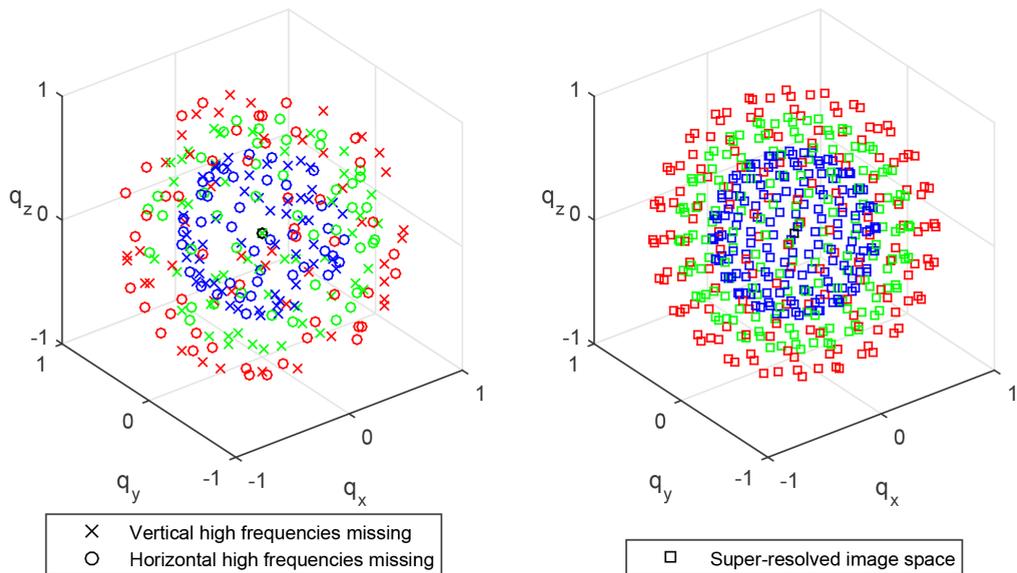


Figure 8.1: Sampling scheme in q -space during acquisition (left) and reconstruction (right). The acquired data have alternating artificial subsampling in vertical/horizontal high frequencies in k -space. All high frequencies for all images are reconstructed. Colors encode the b -value: $\mathcal{B} = \{0, 1000, 2000, 3000\}$ s/mm².

8.2.2.1 Sampling scheme in (k, q) -space

In order to verify the super-resolution reconstruction capability of our holistic reconstruction, we use data of uniquely high resolution from the Human Connectome Project [VESB⁺13, FMS⁺10, SCAG⁺12, XLS⁺12, SJX⁺13, GSW⁺13, JBB⁺12, Fis12], assuming it to be the ground truth underlying image data, and simulate a low-resolution k -space sampling of these ground truth images. In order to leverage complementarity of data in q -space, we employ a low-resolution (k, q) -space sampling scheme [GSM⁺15] in which high resolution components are left out alternatingly in vertical or horizontal image directions for different q -space coordinates. The q -space coordinates and the respective alternating vertical/horizontal k -space subsampling are shown in Figure 8.1, left. Both acquisition and reconstruction (see next paragraph) use the set of b -values $\mathcal{B} = \{0, 1000, 2000, 3000\}$ s/mm².

8.2.2.2 Super-resolution sampling scheme in reconstruction space

While data are artificially subsampled in k -space for the experiments, the reconstruction space is discretized such that the original high image resolution is reconstructed. While 270 q -space coordinates are sampled (Figure 8.1, left), 486 are reconstructed (Figure 8.1,

right). This scheme achieves a super-resolution reconstruction in image and diffusion space.

8.2.2.3 Regularization

We will regularize several images of the type $U \in \mathbb{H}^2(\mathbb{R}^3 \times S^2)$, namely the ODF and the spherical shells in q -space.

The ODF [LWC⁺03] for image r at image location $\mathbf{y} \in \Omega$ and direction $\mathbf{n} \in S^2$ can be calculated as

$$\text{ODF}(r)(\mathbf{y}, \mathbf{n}) = \frac{1}{Z_\kappa} \int_0^\infty (\mathcal{F}_{4,5,6}r)(\mathbf{y}, p\mathbf{n}) p^\kappa dp \quad (8.10)$$

with the usual choice $\kappa = 2$, where Z_κ is a normalization constant and $\mathcal{F}_{4,5,6}$ is the Fourier transform along the diffusion dimensions four to six that calculates the diffusion propagator from q -space data in an idealized setting [Ste65].

Let G_b be the linear operator that extracts a spherical q -space shell at a given b -value (diffusion weighting) from r :

$$(G_b(r))(\mathbf{y}, \mathbf{n}) = r(\mathbf{y}, \sqrt{b}\mathbf{n}). \quad (8.11)$$

In a proof-of-concept holistic reconstruction, the shells and the ODFs are regularized in the uncoupled space $\mathbb{R}^3 \times S^2$ of positions and orientations as follows:

$$\begin{aligned} R(r) = & \sum_{b \in \mathcal{B}} \int_{\mathbb{R}^3 \times S^2} \alpha_1 \|\nabla_{\mathbf{y}} G_b(r)(\mathbf{y}, \mathbf{n})\|^2 \\ & - \alpha_2 \langle G_b(r)(\mathbf{y}, \mathbf{n}), \Delta_{S^2} G_b(r)(\mathbf{y}, \mathbf{n}) \rangle + \alpha_3 |\Delta_{S^2} G_b(r)(\mathbf{y}, \mathbf{n})|^2 d\mathbf{y} d\sigma(\mathbf{n}) \\ & + \int_{\mathbb{R}^3 \times S^2} \alpha_4 \|\nabla_{\mathbf{y}} \text{ODF}(r)(\mathbf{y}, \mathbf{n})\|^2 \\ & - \alpha_5 \langle \text{ODF}(r)(\mathbf{y}, \mathbf{n}), \Delta_{S^2} \text{ODF}(r)(\mathbf{y}, \mathbf{n}) \rangle + \alpha_6 |\Delta_{S^2} \text{ODF}(r)(\mathbf{y}, \mathbf{n})|^2 d\mathbf{y} d\sigma(\mathbf{n}), \end{aligned} \quad (8.12)$$

where \mathcal{B} is the set of reconstructed b -values, the α_i are regularization parameters, σ is the usual surface measure on S^2 , Δ_{S^2} is the Laplace–Beltrami operator on the sphere and the negative inner products correspond to first-order regularization according to $\int -\langle U, \Delta U \rangle = \int \|\nabla U\|^2$ (*i.e.* Green’s identity with vanishing boundary conditions as we assume our functions U to vanish at the boundary).

Defining appropriate inner products on the space $\mathbb{H}^2(\mathbb{R}^3 \times S^2) \ni U, V$ and on $\mathbb{H}^1(\mathbb{R}^3 \times S^2, \mathbb{R}^3) \ni \nabla_{\mathbf{y}} U, \nabla_{\mathbf{y}} V$ as

$$\langle U, V \rangle = \int_{\mathbb{R}^3 \times S^2} U(\mathbf{y}, \mathbf{n}) V(\mathbf{y}, \mathbf{n}) d\mathbf{y} d\sigma(\mathbf{n}), \quad (8.13)$$

$$\langle \nabla_{\mathbf{y}} U, \nabla_{\mathbf{y}} V \rangle = \sum_{i \in \{1,2,3\}} \int_{\mathbb{R}^3 \times S^2} (\nabla_{\mathbf{y}} U(\mathbf{y}, \mathbf{n}))_i (\nabla_{\mathbf{y}} V(\mathbf{y}, \mathbf{n}))_i d\mathbf{y} d\sigma(\mathbf{n}), \quad (8.14)$$

and using the induced norms, we can rewrite the problem (8.9),(8.12) as follows:

$$\begin{aligned} & \min_{r, \varphi} \frac{1}{2} \|T(r, \varphi) - \mathbf{d}\|^2 \\ & + \sum_{b \in \mathcal{B}} \alpha_1 \|\nabla_{\mathbf{y}} G_b(r)\|^2 - \alpha_2 \langle G_b(r), \Delta_{S^2} G_b(r) \rangle + \alpha_3 \|\Delta_{S^2} G_b(r)\|^2 \\ & + \alpha_4 \|\nabla_{\mathbf{y}} \text{ODF}(r)\|^2 - \alpha_5 \langle \text{ODF}(r), \Delta_{S^2} \text{ODF}(r) \rangle + \alpha_6 \|\Delta_{S^2} \text{ODF}(r)\|^2. \end{aligned} \quad (8.15)$$

8.2.2.4 Reformulations

To obtain a convenient min-max form with simpler expressions within the norms, we shall use the identity:

$$\|\hat{x}\|^2 = \sup_{\hat{y}} \langle \hat{x}, \hat{y} \rangle - \frac{1}{4} \|\hat{y}\|^2, \quad (8.16)$$

obtained by taking the convex biconjugate and completing the square. This reformulation introduces dual variables \hat{y} .

8.2.2.5 Optimization procedure

Our optimization problem (8.15) can be rewritten as a min-max problem of the form

$$\min_x \max_y G(x) + \langle K(x), y \rangle - F^*(y) \quad (8.17)$$

with convex G , F^* and a nonlinear K , which can be solved with the modified primal-dual hybrid gradient method for nonlinear K [PCBC09, CP11, Val14]:

$$x^{i+1} := (I + \tau \partial G)^{-1}(x^i - \tau [\nabla K(x^i)]^* y^i), \quad (8.18a)$$

$$x_\omega^{i+1} := x^{i+1} + \omega(x^{i+1} - x^i), \quad (8.18b)$$

$$y^{i+1} := (I + \sigma \partial F^*)^{-1}(y^i + \sigma K(x_\omega^{i+1})), \quad (8.18c)$$

where ∂f represents the subdifferential of a function f , defined as

$$\partial f(x_0) = \{v \mid f(x) - f(x_0) \geq \langle v, x - x_0 \rangle \forall x \in \text{dom} f\}, \quad (8.19)$$

and $(I + \lambda \partial f)^{-1}$ is the resolvent of the subdifferential, corresponding to the proximal operator [PB14]:

$$(I + \lambda \partial f)^{-1} x = \text{prox}_{\lambda f}(x) = \arg \min_z f(z) + \frac{1}{2\lambda} \|x - z\|^2. \quad (8.20)$$

The algorithm (8.18) has been applied [Val14] with the operator $T(r, \varphi)$ to non-diffusion MRI, and with another operator to dMRI. The author announces combining $T(r, \varphi)$ with direct reconstruction of the diffusion tensor in a future study, while we present an application of $T(r, \varphi)$ to reconstruction in image×diffusion space.

By rewriting all five norms in our problem (8.15) using the identity (8.16), we obtain the min-max form

$$\begin{aligned}
\min_{r, \varphi} \max_{\lambda, (\zeta_b)_{b \in \mathcal{B}}, (\eta_b)_{b \in \mathcal{B}}, \xi, \nu} & \langle T(r, \varphi), \lambda \rangle - \langle d, \lambda \rangle - \frac{1}{2} \|\lambda\|^2 \\
& + \sum_{b \in \mathcal{B}} \alpha_1 \left(\langle \nabla_{\mathbf{y}} G_b(r), \zeta_b \rangle - \frac{1}{4} \|\zeta_b\|^2 \right) \\
& - \alpha_2 \langle G_b(r), \Delta_{S^2} G_b(r) \rangle + \alpha_3 \left(\langle \Delta_{S^2} G_b(r), \eta_b \rangle - \frac{1}{4} \|\eta_b\|^2 \right) \\
& + \alpha_4 \left(\langle \nabla_{\mathbf{y}} \text{ODF}(r), \xi \rangle - \frac{1}{4} \|\xi\|^2 \right) \\
& - \alpha_5 \langle \text{ODF}(r), \Delta_{S^2} \text{ODF}(r) \rangle + \alpha_6 \left(\langle \Delta_{S^2} \text{ODF}(r), \nu \rangle - \frac{1}{4} \|\nu\|^2 \right).
\end{aligned} \tag{8.21}$$

The primal variables are $x = (r, \varphi)$ and the dual ones are $y = (\lambda, (\zeta_b)_{b \in \mathcal{B}}, (\eta_b)_{b \in \mathcal{B}}, \xi, \nu)$, where for example η_b denotes the dual variable associated to $\|\Delta_{S^2} G_b(r)\|^2$. This can be regrouped into the standard form (8.17) as follows:

$$\begin{aligned}
G(x) &= \sum_{b \in \mathcal{B}} -\alpha_2 \langle G_b(r), \Delta_{S^2} G_b(r) \rangle - \alpha_5 \langle \text{ODF}(r), \Delta_{S^2} \text{ODF}(r) \rangle, \\
\langle K(x), y \rangle &= \langle T(r, \varphi), \lambda \rangle + \sum_{b \in \mathcal{B}} \alpha_1 \langle \nabla_{\mathbf{y}} G_b(r), \zeta_b \rangle + \alpha_3 \langle \Delta_{S^2} G_b(r), \eta_b \rangle \\
&+ \alpha_4 \langle \nabla_{\mathbf{y}} \text{ODF}(r), \xi \rangle + \alpha_6 \langle \Delta_{S^2} \text{ODF}(r), \nu \rangle, \\
\pm F^*(y) &= \pm \langle d, \lambda \rangle \pm \frac{1}{2} \|\lambda\|^2 \\
&\pm \frac{1}{4} \left(\sum_{b \in \mathcal{B}} \alpha_1 \|\zeta_b\|^2 + \alpha_3 \|\eta_b\|^2 + \alpha_4 \|\xi\|^2 + \alpha_6 \|\nu\|^2 \right).
\end{aligned} \tag{8.22}$$

For the implementation of algorithm (8.18), we calculate the proximal operators [PB14]:

$$(I + \tau \partial G)^{-1} x = (I + \tau(Q + Q^*))^{-1} x, \tag{8.23}$$

$$Q = \sum_{b \in \mathcal{B}} G_b^* \Delta_{S^2} G_b + \text{ODF}^* \Delta_{S^2} \text{ODF}, \tag{8.24}$$

$$(I + \sigma \partial F^*)^{-1} y = \begin{pmatrix} (\lambda - \sigma d)/(\sigma + 1) \\ (\zeta_b/(1 + \alpha_1 \sigma/2))_{b \in \mathcal{B}} \\ (\eta_b/(1 + \alpha_3 \sigma/2))_{b \in \mathcal{B}} \\ \xi/(1 + \alpha_4 \sigma/2) \\ \nu/(1 + \alpha_6 \sigma/2) \end{pmatrix}. \tag{8.25}$$

Calculating $[\nabla K(x^i)]^*$ (8.18) for the nonlinear part $T(r, \varphi)$ (8.22) yields

$$[\nabla T(r, \varphi)]^* = (S\mathcal{F}_{1,2}[\nabla C(r, \varphi)])^* = [\nabla C(r, \varphi)]^* \mathcal{F}_{1,2}^* S^*, \quad (8.26)$$

$$[\nabla C(r, \varphi)]^* \hat{\lambda} = \begin{pmatrix} \Re(\hat{\lambda}) \cos(\varphi) + \Im(\hat{\lambda}) \sin(\varphi) \\ r(\Im(\hat{\lambda}) \cos(\varphi) - \Re(\hat{\lambda}) \sin(\varphi)) \end{pmatrix}. \quad (8.27)$$

8.2.2.6 Unbounded ODF operator

When writing out the Fourier transform $\mathcal{F}_{4,5,6}$ over $Q \in \mathbb{R}^3$, the ODF (8.10) contains the diverging term $\exp(-i(p \mathbf{n}, Q))p^2$. Thus, the ODF operator is unbounded. Since an adjoint is required for the algorithm (8.18), the operator can be made bounded in the infinite-dimensional setting by including a Gaussian damping factor $\exp(-p^2/\zeta^2)$ as a mollifier. The operator bound of the discrete operator depends on the discretization, and in our discretization scheme no mollifier was needed in practice.

8.2.2.7 Implementation details

The operators $\mathcal{F}_{1,2}$, S (8.6), ODF (8.10), G_b (8.11), ∇_y and Δ_{S^2} are linear. In the implementation, the spaces in which acquisition, regularization and reconstruction take place are discretized and thus the operators can be written as matrices. We obtain these matrices explicitly. Where not evident, an operator matrix is computed by applying the operator to all standard basis vectors of the discretized space, yielding the columns of the matrix. For pointwise operators, we compute and store repeating coefficients only once. When computing $[\nabla K(\cdot)]^*$ and $K(\cdot)$ in the algorithm (8.18), having the operator matrices explicitly has the advantages of rapid computation by matrix multiplication and easy computation of the adjoint operators. Besides, in the discretized setting, the ODF operator is not unbounded anymore and thus has an adjoint, as required by the algorithm. The norm $\|[\nabla K(\cdot)]^*\|$ of the operator $[\nabla K(\cdot)]^*$ explodes as the discretization becomes finer, but in our discretization settings there was no need to include a Gaussian mollifier in (8.10).

8.3 Results

Figure 8.2 shows the high-resolution ‘‘ground truth’’ image data from the Human Connectome Project (Figure 8.2, left) alongside the results of two reconstruction methods applied to the same data that has been artificially subsampled according to the sampling scheme in (k, q) -space described in section 8.2.2.1 and illustrated in Figure 8.1, left. This artificial subsampling procedure emulates a clinical setting where resolution is considerably lower than in the Human Connectome Project, and enables a comparison to this

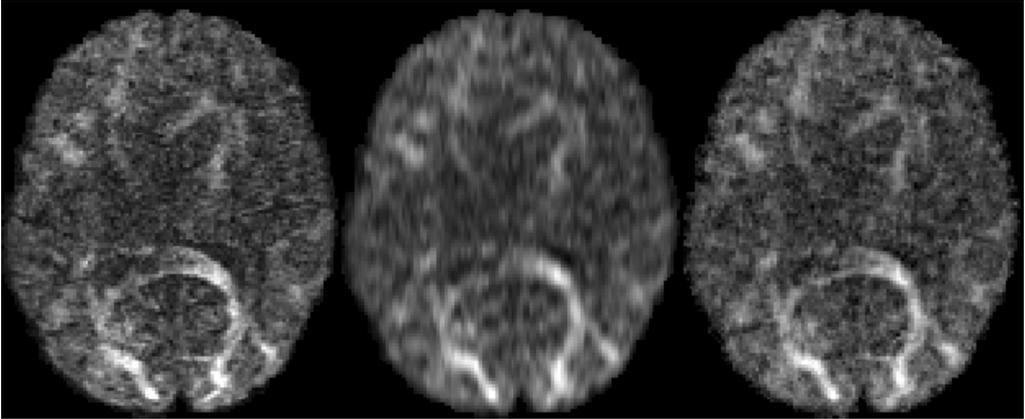


Figure 8.2: High-resolution ground truth (left), standard reconstruction (middle), holistic super-resolution reconstruction (right).

exceptionally high-resolution ground truth data. The two compared reconstruction methods are standard reconstruction ($\mathcal{F}_{1,2}$ -transformed subsampled data; Figure 8.2, middle) and holistic image reconstruction (as described above, with super-resolution sampling as in Figure 8.1, right; results in Figure 8.2, right).

The employed parameters were $\alpha_1 = 0.3$, $\alpha_2 = 0.1$, $\alpha_3 = 0.1$, $\alpha_4 = 0.01$, $\alpha_5 = 0.3$, $\alpha_6 = 0.01$.

Holistic image reconstruction demonstrates considerably more detail than standard reconstruction. While standard reconstruction results have a visibly lower resolution, holistic reconstruction retrieves details that are present in the ground truth data due to its super-resolution scheme and regularization in image and diffusion space.

8.4 Conclusion

Among the numerous advanced dMRI reconstruction methods existing in literature, many methods perform denoising, missing data reconstruction (q -space compressed sensing), enhancement, etc. as an intermediate post-processing step *after* image-space reconstruction. However, standard-reconstructed images can contain artifacts, intensity bias (*e.g.* Rician or more complicated), and irretrievably discard some parts of information present in the raw k -space data. Imposing data consistency in reconstructed image space can lead to these errors being propagated on into subsequent data processing steps, and/or introduce less tractable bias-correction terms. There is strong evidence that one-step pipelines are better than multi-step pipelines due to information loss in intermediate steps [GDS⁺15]. Particularly, imposing data consistency on the original raw data

in k -space yields improved results compared to multi-step processing [GSM⁺15]. The holistic reconstruction framework presented herein allows imposing data consistency in the original data acquisition space, while also including regularization in several spaces (such as (y, q) -space and “ (y, ODF) -space”), and reconstructing into an arbitrary space, including super-resolution reconstruction sampling.

Super-resolution methods are beneficial for dMRI due to their capability to exceed hardware limitations on resolution. In the presented holistic reconstruction framework, super-resolution is performed in image space *and* diffusion space simultaneously, cf. Figure 8.1. At the same time, data consistency in the original space and regularizations in additional spaces are incorporated in a straightforward manner.

Many competing regularizers in different spaces exist in recent literature. Each of them incorporates certain assumptions and improves data quality at certain intermediate regularization strengths. Regularizations in different spaces can be combined into one procedure (including true data consistency and super-resolution) using holistic image reconstruction.

Reconstruction can be performed jointly with motion and distortion correction [TTS⁺14] in the future. Finally, our choice of priors in (8.15) was based on isotropic Laplacians over the spatial and angular part, and as such defined on $\mathbb{R}^3 \times S^2$. Including anisotropies and alignment modeling in a crossing-preserving way via the *coupled* space $\mathbb{R}^3 \times S^2$, as was done in Chapter 6, is expected to give better results in future work.

Concluding Remarks

I would like to end with a number of concluding remarks for each of the chapters, summarizing and reflecting on what I consider to be the most important results of my research and adding a personal note here and there.

The main achievement of Chapter 3 was that we have presented exact solutions to hypo-elliptic diffusion and convection-diffusion equations on $SE(3)$. It was originally set out as one of the milestones for this PhD-project, as a logical but challenging next step after extensive previous research on these PDEs on $SE(2)$. Although the derivation of the current solution contains mathematical ingredients that might be used elsewhere, it extends just one of three approaches that were presented for the $SE(2)$ case. It would be interesting to see if the other two approaches (resulting in a rapidly decaying series, or an explicit computation of such rapidly decaying series towards a single term) also have an extension in $SE(3)$.

Chapter 4 can be seen as the aggregation of different pieces of research that was done in the context of approximations of the kernels to the hypo-elliptic diffusion equation. The discussion on certain symmetries of these kernels was started after I discovered the non-symmetric behavior of kernels that we used so far. The link between the kernels and sub-Riemannian distances was made more explicit in later work, which is why I thought the two parts could nicely be combined into one chapter. I was happy to see that the kernels we derived were implemented ([MSG⁺16a]) in the practical neuroimage software package Dipy and I am curious to see if this is going to be picked up by others. Also the application of perceptual organization of local orientations as presented in [BCP17] can have an interesting future, perhaps in a combination of dMRI and MR Angiography. Finally, the stochastic Monte Carlo approach in this chapter provided yet another viewpoint on the heat kernels and I found it quite insightful to see the various approaches lined up.

Chapter 5 provides the mathematical background that allows us to solve a variety

of 3D shortest path problems. We can now start looking for new applications, either in image analysis or in e.g. robotics, and at the same time explore if better results can be obtained using new metrics. These metrics can be based on data-adaptive frames, or perhaps even make use of a machine learning setting to determine model parameters. The paper I am currently working on in collaboration with Jean-Marie Mirebeau, that is accompanied by the distribution of the fast marching code, should also enable others to implement their ideas on these optimal path problems.

In Chapter 6, we have shown the benefit of hypo-elliptic (linear) diffusion for tractography on dMRI data. A logical next step would be to apply nonlinear, edge-preserving diffusion on positions and orientations. Based on the considerable advantage of nonlinear enhancements over linear diffusion in regular image analysis, I expect to see much better results for processing of dMRI data and subsequent tractography.

In Chapter 7, we have demonstrated how the curvature penalized shortest path framework can be applied in dMRI data. The results in this chapter concerning the proposed connectivity measures are promising, but also leave room for improvement. In future research, improvement could be found in extending the definition of connectivity between points to connectivity between regions, or by combining global with local connectivity measures. During the three months of my short postdoc, I hope to make some progress with this application.

In Chapter 8, we introduced a framework in which multiple concepts and regularizations could be combined into a single optimization problem, for super-resolution reconstruction of dMRI. The optimization problem was solved numerically using a primal-dual gradient method. The advantage of this framework, being the fact that several regularization types can be included in a single optimization, also has the downside that it is not entirely clear what the important regularization terms are. After more insight in this matter is obtained, it would be interesting to see potential benefits of anisotropic priors.

A

Appendix to Chapter 3

A.1 Expansion of spheroidal wave functions

In this section we show how to obtain the eigenfunctions of the spheroidal wave equation (3.36). First note that when $\rho = r = 0$, corresponding to $\boldsymbol{\omega} = \mathbf{0}$, the spheroidal wave equation reduces to the Legendre differential equation, that has eigenvalues $-l(l+1)$ and eigenfunctions P_l^m , associated Legendre functions with $l \in \mathbb{N}_0$, $m \in \mathbb{Z}$, $|m| \leq l$. We immediately include a normalization factor in the definition of these functions:

$$P_l^m(x) = N^{l,m} (1-x^2)^{|m|/2} \frac{d^{|m|} P_l(x)}{dx^{|m|}}, \quad P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} [(x^2-1)^l], \quad -1 \leq x \leq 1. \quad (\text{A.1})$$

The normalization factor $N^{l,m}$ is given by

$$N^{l,m} = \sqrt{\frac{(2l+1)(l-|m|)!}{2(l+|m|)!}}. \quad (\text{A.2})$$

When $\rho > 0$ in Eq. (3.36), several series solutions are possible, but it is customary to use a series of associated Legendre polynomials [Fla57, OLB10]:

$$S_\rho^{l,m}(x) = \sum_{j=0}^{\infty} d_j^{l,m} P_{m+j}^m(x), \quad m \geq 0, \quad (\text{A.3})$$

where $d_j^{l,m} = d_j^{l,m}(\rho)$. For the case $m < 0$ it follows immediately from the property that $P_l^m(x) = P_l^{-m}(x)$ that:

$$d_j^{l,m} = d_j^{l,-m}, \quad S_\rho^{l,m}(x) = S_\rho^{l,-m}(x) = \sum_{j=0}^{\infty} d_j^{l,|m|} P_{|m|+j}^m(x), \quad m \in \mathbb{Z}. \quad (\text{A.4})$$

We use the following identity to normalize the $S_\rho^{l,m}$:

$$\int_{-1}^1 S_\rho^{l,m}(x) S_\rho^{l',m}(x) dx = \delta_{ll'} \sum_{j=0}^{\infty} |d_j^{l,m}|^2 =: \delta_{ll'} \|\mathbf{d}^{l,m}\|^2, \quad \mathbf{d}^{l,m} := (d_j^{l,m})_{j=0}^{\infty}. \quad (\text{A.5})$$

Our solutions are of the form $S_\rho^m(x) = \sum_{j=0}^{\infty} d_j^m(\rho) P_{m+j}^m(x)$, where in this appendix we only consider $m \geq 0$. For shortness, we will omit the dependence on ρ of the coefficients. It will follow later that for every m a countable number of solutions exist for S_ρ^m . For m fixed, substitution of the series in the differential equation (3.36) gives the following identity:

$$\sum_{j=0}^{\infty} (-\rho^2 x^2 + \tilde{\lambda}_\rho - (m+j)(m+j+1)) d_j^m P_{m+j}^m(x) = 0, \quad \text{i.e.,} \quad (\text{A.6})$$

$$-\rho^2 \sum_{j=0}^{\infty} d_j^m x^2 P_{m+j}^m(x) + \sum_{j=0}^{\infty} (\tilde{\lambda}_\rho - (m+j)(m+j+1)) d_j^m P_{m+j}^m(x) = 0. \quad (\text{A.7})$$

By applying the identity given in (3.91), we can expand the $x^2 P_{m+j}^m$ term in Legendre polynomials P_{m+j-2}^m , P_{m+j}^m and P_{m+j+2}^m . By substituting Eq. (3.91) in Eq. (A.7) and by equating coefficients of $P_{m+j}^m(x)$, the following relation for the d 's can be found:

$$\begin{aligned} \rho^2 \frac{N^{m+j+2,m}}{N^{m+j,m}} \alpha^{m+j+2,m} d_{j+2}^m + (\rho^2 \eta^{m+j,m} + (m+j)(m+r+1)) d_j^m \\ + \rho^2 \frac{N^{m+j-2,m}}{N^{m+j,m}} \zeta^{m+j-2,m} d_{j-2}^m = \tilde{\lambda}_\rho d_j^m. \end{aligned} \quad (\text{A.8})$$

In matrix form, this equation can be written as

$$(\rho^2 \mathbf{M}_1^m + \mathbf{\Lambda}^m) \mathbf{d}^m = \tilde{\lambda}_\rho \mathbf{d}^m, \quad \mathbf{d}^m = (d_0^m, d_1^m, \dots)^T, \quad (\text{A.9})$$

with \mathbf{M}_1^m as in (3.92) and $\mathbf{\Lambda}^m$ as in (3.97).

Then the eigenvalues of the matrix on the left are spheroidal eigenvalues, that we denote with $\tilde{\lambda}_\rho^{l,m}$, $l \geq |m|$, such that $\tilde{\lambda}_\rho^{m,m} < \tilde{\lambda}_\rho^{m+1,m} < \tilde{\lambda}_\rho^{m+2,m} < \dots$. The corresponding eigenvectors form the constant vectors $\mathbf{d}^{l,m}$ and thereby the functions $S_\rho^{l,m}(x)$ are determined up to normalization. It follows from the form of the matrix in (A.9) that either the even or the odd coefficients of $\mathbf{d}^{l,m}$ are 0, resulting in only even and odd functions $S_\rho^{l,m}(x)$. Recall that in our case $\rho = \sqrt{\frac{D_{33}}{D_{44}}} r > 0$ and the eigenvalues corresponding to the eigenfunctions $\Phi_{l,m}^\omega$ are $\lambda_r^{l,m} = -D_{44} \tilde{\lambda}_\rho^{l,m}$, and thereby

$$(r^2 D_{33} \mathbf{M}_1^m + D_{44} \mathbf{\Lambda}^m) \mathbf{d}^{l,m} = -\lambda_r^{l,m} \mathbf{d}^{l,m}. \quad (\text{A.10})$$

We conclude that the spheroidal wave functions are given by

$$S_\rho^{l,m}(x) = \sum_{j=0}^{\infty} d_j^{l,|m|}(\rho) P_{|m|+j}^m(x), \quad l \in \mathbb{N}_0, m \in \mathbb{Z}, |m| \leq l, \quad (\text{A.11})$$

where the vectors $\mathbf{d}^{l,m}(\rho)$ are the solutions to Eq. (A.10).

A.2 Generalized spheroidal wave functions

Similar to the case of the spheroidal wave functions, we can derive eigenfunctions of the generalized spheroidal wave equation by substituting $GS_\rho^m(x) = \sum_{j=0}^{\infty} c_j^m(\rho) P_{m+j}^m(x)$ in Eq. (3.61). For shortness, we omit the dependence on ρ of the coefficients and we only consider the case $m \geq 0$. This yields

$$-i\rho \sum_{j=0}^{\infty} c_j^m x P_{m+j}^m(x) + \sum_{j=0}^{\infty} (\tilde{\lambda}_\rho - (m+j)(m+j+1)) c_j^m P_{m+j}^m(x) = 0. \quad (\text{A.12})$$

In our case $\rho = \frac{r}{D_{44}} > 0$ and the eigenvalues $\lambda_r^{l,m}$ of the eigenfunctions $\Psi_{l,m}^\omega$ are given by $\lambda_r^{l,m} = -D_{44} \tilde{\lambda}_\rho^{l,m}$. Now applying (3.89) once to rewrite the term $x P_{m+j}^m(x)$, we get

$$\begin{aligned} -i\rho \sum_{j=0}^{\infty} c_j^m \frac{N^{m+j}}{N^{m+j-1}} \nu^{m+j,m} P_{m+j-1}^m(x) + \frac{N^{m+j}}{N^{m+j+1}} c_j^m \xi^{m+j,m} P_{m+j+1}^m(x) \\ + \sum_{j=0}^{\infty} (\tilde{\lambda}_\rho - (m+j)(m+j+1)) c_j^m P_{m+j}^m(x) = 0. \end{aligned} \quad (\text{A.13})$$

Again, equating coefficients of P_{m+j}^m , we get

$$i\rho \nu^{m+j+1,m} c_{j+1}^m + i\rho \xi^{m+j-1,m} c_{j-1}^m + (m+j)(m+j+1) c_j^m = \tilde{\lambda}_\rho c_j^m, \quad j \in \mathbb{N}_0. \quad (\text{A.14})$$

In matrix form:

$$(i\rho \mathbf{M}_2^m + \mathbf{\Lambda}^m) \mathbf{c}^m = \tilde{\lambda}_\rho \mathbf{c}^m, \quad \mathbf{c}^m = (c_0^m, c_1^m \dots)^T, \quad (\text{A.15})$$

with \mathbf{M}_2^m as in (3.103) and $\mathbf{\Lambda}^m$ as in (3.97).

For each fixed m , eigenvalues $\tilde{\lambda}_\rho$ can again be numbered as $\tilde{\lambda}_\rho^{l,m}$, $l \geq |m|$, with eigenvectors \mathbf{c}^m . This determines the eigenfunctions $GS_\rho^{l,m}$ up to a normalization constant:

$$GS_\rho^{l,m}(x) = \sum_{j=0}^{\infty} c_j^{l,|m|}(\rho) P_{|m|+j}^m(x), \quad l \in \mathbb{N}_0, m \in \mathbb{Z}, |m| \leq l, \quad (\text{A.16})$$

where the vectors $\mathbf{c}^{l,m}(\rho)$ are the solutions to Eq. (A.15).

A.3 Proof of Theorem 3.3.6

We now give a proof for Theorem 3.3.6, that stated the relation between time-integration of the diffusion and convection-diffusion kernels and the repeated convolution of resolvent kernels.

Proof. When $T \sim \Gamma(k, \alpha)$, we can also write $T = \sum_{j=1}^k T_k$ with i.i.d. $T_k \sim \text{Exp}(\alpha)$ and corresponding pdf $P_\alpha(t) := \alpha e^{-\alpha t}$. When $k = 1$, we have seen that $R_\alpha^i = -\alpha(Q_i - \alpha I)^{-1} \delta_{0, \mathbf{e}_z}$. When $k > 1$, we have

$$\begin{aligned}
 \int_0^\infty K_t^i \Gamma(t; k, \alpha) dt &= \int_0^\infty e^{tQ_i} \delta_{(\mathbf{0}, \mathbf{e}_z)} (P_\alpha *_{\mathbb{R}^+}^{(k-1)} P_\alpha)(t) dt \\
 &= [\alpha [\mathcal{L}(t \mapsto e^{tQ_i} \delta_{(\mathbf{0}, \mathbf{e}_z)})] (\alpha)]^k \\
 &= \alpha^k [(\alpha I - Q_i)^{-k}] \delta_{(\mathbf{0}, \mathbf{e}_z)} = R_\alpha^i *_{\mathbb{R}^3 \times S^2}^{(k-1)} R_\alpha^i,
 \end{aligned} \tag{A.17}$$

with \mathcal{L} the Laplace transform on \mathbb{R}^+ and $*_{\mathbb{R}^+}$ the convolution on \mathbb{R}^+ . In the second step we used the property that $\mathcal{L}(f *_{\mathbb{R}^+} g) = (\mathcal{L}f)(\mathcal{L}g)$.

Regarding the singularities of these kernels, we recall that

$$\begin{aligned}
 \hat{K}_t^1(\boldsymbol{\omega}, \mathbf{n}) &= \sum_{l=0}^\infty \sum_{m=-l}^l \overline{\Phi_{l,m}^\omega(\mathbf{e}_z)} \Phi_{l,m}^\omega(\mathbf{n}) e^{\lambda_r^{l,m} t}, & \lambda_r^{l,m} &= -D_{44}l(l+1) + \mathcal{O}(r^2), \\
 \hat{K}_t^2(\boldsymbol{\omega}, \mathbf{n}) &= \sum_{l=0}^\infty \sum_{m=-l}^l \overline{\Psi_{l,m}^\omega(\mathbf{e}_z)} \Psi_{l,m}^\omega(\mathbf{n}) e^{\lambda_r^{l,m} t}, & \lambda_r^{l,m} &= -D_{44}l(l+1) + \mathcal{O}(r).
 \end{aligned} \tag{A.18}$$

All eigenfunctions $\Phi_{l,m}^\omega, \Psi_{l,m}^\omega$ are bounded on the compact set S^2 . Consider the case $i = 1$, then

$\mathcal{F}_{\mathbb{R}^3} \left[R_\alpha^1 *_{\mathbb{R}^3 \times S^2}^{(k-1)} R_\alpha^1 \right] (\boldsymbol{\omega}, \mathbf{n})$ is \mathbb{L}_1 -integrable w.r.t $\boldsymbol{\omega}$ if

$$\int_0^\infty \left(\frac{\alpha}{l(l+1)D_{44} + \mathcal{O}(r^2) + \alpha} \right)^k r^2 dr < \infty. \tag{A.19}$$

This implies $k > \frac{3}{2}$. For the case $i = 2$, it follows analogously that $k > 3$. The fact that this implies that no singularities occur when these conditions for k are satisfied follows from [Rud06, Thm. 7.5].

□

A.4 Equivalent solutions via the Fourier transform on $SE(3)$

In this appendix, we show how to obtain the solutions of the diffusion and convection-diffusion equations using the Fourier transform on $SE(3)$, rather than a Fourier transform in only the spatial coordinates. We write $G = SE(3)$ for shortness. Recall from Section 2.1 that we denote the rigid body motions with $g = (\mathbf{x}, \mathbf{R}) \in SE(3) = \mathbb{R}^3 \rtimes SO(3)$. Now G is a unimodular Lie group (of type I) with (left- and right-invariant) Haar measure $dg = d\mathbf{x}d\mu_{SO(3)}(\mathbf{R})$ being the product of the Lebesgue measure on \mathbb{R}^3 and the Haar-Measure $\mu_{SO(3)}$ on $SO(3)$. We denote the unitary dual space of G with \hat{G} . Its elements are equivalence classes of unitary, irreducible group representations $\sigma : G \rightarrow B(H_\sigma)$, where $B(H_\sigma) = \{A : H_\sigma \rightarrow H_\sigma \mid A \text{ linear and } \text{trace}(A^*A) < \infty\}$ denotes the space of

bounded linear trace-class operators on a Hilbert space H_σ (on which each operator σ_g acts). In the dual space \hat{G} we identify elements via the following equivalence relation:

$$\sigma_1 \sim \sigma_2 \Leftrightarrow \text{there exists a unitary operator } U, \text{ s.t. } \sigma_1 = U \circ \sigma_2 \circ U^{-1}.$$

Then for all $f \in \mathbb{L}_1(G) \cap \mathbb{L}_2(G)$ the Fourier transform $\mathcal{F}_G f$ is given by

$$\hat{f}(\sigma) = (\mathcal{F}_G(f))(\sigma) = \int_G f(g) \sigma_{g^{-1}} dg \in B(H_\sigma), \text{ for all } \sigma \in \hat{G}, \quad (\text{A.20})$$

where \mathcal{F}_G maps $\mathbb{L}_2(G)$ unitarily onto the direct integral space $\int_{\hat{G}}^{\oplus} B(H_\sigma) d\nu(\sigma)$ with ν the Plancherel measure on \hat{G} . For more details see [Füh05, Sug90, Fol94]. One has the inversion formula:

$$f(g) = (\mathcal{F}_G^{-1} \mathcal{F}_G f)(g) = \int_{\hat{G}} \text{trace}\{(\mathcal{F}_G f)(\sigma) \sigma_g\} d\nu(\sigma). \quad (\text{A.21})$$

In our Lie group case of $SE(3)$ we identify all irreducible representations $\sigma^{p,s}$ having non-zero dual measure with the pair $(p, s) \in \mathbb{R}^+ \times \mathbb{Z}$. This identification is commonly applied, see e.g. [CK00]. All unitary irreducible representations (UIR's) of G , up to equivalence, with non-zero Plancherel measure are given by [Mac49, Sug90]:

$$\begin{aligned} \sigma = \sigma^{p,s} : SE(3) &\rightarrow B(\mathbb{L}_2(pS^2)), \quad p > 0, s \in \mathbb{Z}, \text{ given by} \\ (\sigma_{(\mathbf{x}, \mathbf{R})}^{p,s} \phi)(\mathbf{u}) &= e^{-i\mathbf{u} \cdot \mathbf{x}} \phi(\mathbf{R}^{-1} \mathbf{u}) \Delta_s \left(\mathbf{R}_{\frac{\mathbf{u}}{p}}^{-1} \mathbf{R} \mathbf{R}_{\frac{\mathbf{u}}{p}} \right), \quad \mathbf{u} \in pS^2, \phi \in \mathbb{L}_2(pS^2), \end{aligned} \quad (\text{A.22})$$

where Δ_s is a representation of $SO(2)$ (or rather of the stabilizing subgroup $\text{stab}(\mathbf{e}_z) \subset SO(3)$ isomorphic to $SO(2)$) producing a scalar. In (A.22) $\mathbf{R}_{\frac{\mathbf{u}}{p}}$ denotes a rotation that maps \mathbf{e}_z onto $\frac{\mathbf{u}}{p}$. So direct computation

$$\mathbf{R}_{\frac{\mathbf{u}}{p}}^{-1} \mathbf{R} \mathbf{R}_{\frac{\mathbf{u}}{p}} \mathbf{e}_z = \mathbf{R}_{\frac{\mathbf{u}}{p}}^{-1} \mathbf{R} \mathbf{R}^{-1} \left(\frac{\mathbf{u}}{p} \right) = \mathbf{e}_z,$$

shows us that it is a rotation around the z -axis, say about angle ϕ . This yields character $\Delta_s(\mathbf{R}_{\frac{\mathbf{u}}{p}}^{-1} \mathbf{R} \mathbf{R}_{\frac{\mathbf{u}}{p}}) = e^{-is\phi}$, for details see [CK00, ch.10.6].

Furthermore, the dual measure ν can be identified with the Lebesgue measure on pS^2 and we have

$$d\nu(\sigma^{p,s}) \equiv p^2 dp \text{ for all } p > 0, s \in \mathbb{Z}.$$

The matrix elements of $\mathcal{F}_G f$ w.r.t. an orthonormal basis of modified spherical harmonics $\{Y_s^{l,m}(p^{-1}\cdot)\}$, with $|m|, |s| \leq l$, for $\mathbb{L}_2(pS^2)$ are given by

$$\hat{f}_{l,m,l',m'}^{p,s} := \int_G f(g) (Y_s^{l,m}(p^{-1}\cdot), \sigma_{g^{-1}}^{p,s} Y_s^{l',m'}(p^{-1}\cdot))_{\mathbb{L}_2(pS^2)} dg. \quad (\text{A.23})$$

For an explicit formula for the modified spherical harmonics, see [CK00], where they are denoted with $h_{m,s}^l$. Moreover, we have inversion formula ([CK00, Eq.10.46]):

$$f(g) = \frac{1}{2\pi^2} \sum_{s \in \mathbb{Z}} \sum_{l'=|s|}^{\infty} \sum_{l=|s|}^{\infty} \sum_{m'=-l}^l \sum_{m=-l}^l \int_0^{\infty} \hat{f}_{l,m,l',m'}^{p,s}(\sigma_g^{p,s})_{l',m',l,m} p^2 dp, \quad (\text{A.24})$$

with matrix coefficients (independent of f) given by

$$(\sigma_g^{p,s})_{l',m',l,m} = (\sigma_g^{p,s} Y_s^{l,m}(p^{-1}\cdot), Y_s^{l',m'}(p^{-1}\cdot))_{\mathbb{L}_2(pS^2)}. \quad (\text{A.25})$$

Note that $\sigma^{p,s}$ is a UIR so we have

$$(\sigma_{g^{-1}}^{p,s})_{l',m',l,m} = \overline{(\sigma_g^{p,s})_{l,m,l',m'}}. \quad (\text{A.26})$$

Now in our position and orientation analysis we must constrain ourselves to a special type of functions $f = \tilde{U} \in \mathbb{L}_2(G)$, namely the ones that have the property that $f(\mathbf{x}, \mathbf{R}) = \tilde{U}(\mathbf{x}, \mathbf{R}) = U(\mathbf{x}, \mathbf{R}e_z)$. Then it follows by [CK00, eq.10.35] and by (A.26) that the only non-zero matrix elements (A.25) have the property that both $m = 0$ and $m' = 0$. This can also be observed in the analytical examples in [CK00, ch:10.10]. This reduces the five-fold sum in (A.24) to a three-fold sum, and the modified spherical harmonics become standard spherical harmonics.

Next we introduce some notation for generators of group representations of $SE(3)$.

Definition A.4.1. *Let $A \in T_e(G)$ denote an element of the Lie algebra of $G = SE(3)$, with unity element $e = (\mathbf{0}, I) \in G$. Then given a group representation $\mathcal{U} : G \rightarrow B(H)$, where $B(H)$ denotes the space of bounded linear operators on some Hilbert space H , we define its generator $d\mathcal{U}(A) : \mathcal{D}_H \rightarrow H$ by*

$$d\mathcal{U}(A)f := \lim_{t \rightarrow 0} t^{-1} (\mathcal{U}_{e^{tA}} - I)f, \quad \text{for all } f \in \mathcal{D}_H,$$

where $\mathcal{D}_H \subset H$ is a domain of sufficiently regular $f \in H$ such that limit $d\mathcal{U}(A)f$ exists and is in H .

Now let $\mathcal{R} : G \rightarrow B(\mathbb{L}_2(G))$ denote the right-regular representation given by $(\mathcal{R}_h f)(g) = f(gh)$, $f \in \mathbb{L}_2(G)$ and $g, h \in G$. Let $\mathcal{A}_i|_g = (L_g)_* A_i$ denote the left-invariant vector fields on G relative to basis $\{A_i\}$ in the Lie-algebra $T_e(G)$. Then the left-invariant vector fields $\{\mathcal{A}_i\}$ are obtained by the generators of the right-regular representation:

$$\mathcal{A}_i f = d\mathcal{R}(A_i)f := \lim_{t \rightarrow 0} t^{-1} (\mathcal{R}_{e^{tA_i}} - I)f, \quad \text{for all } f \in \mathbb{H}(SE(3)). \quad (\text{A.27})$$

Here we note that due to our choice $H = \mathbb{L}_2(SE(3))$ in Definition A.4.1 the domain \mathcal{D}_H of the generator $d\mathcal{R}(A_i)$ becomes the first order Sobolev space $\mathbb{H}(SE(3))$.

In the lemma below we use this identity, together with Definition A.4.1 applied to the UIR's $\sigma^{p,s}$. In fact the lemma allows us to map unbounded generators acting on sufficiently regular $f \in \mathbb{L}_2(G)$ to unbounded generators acting on sufficiently regular $\mathcal{F}_G f \in \mathbb{L}_2(\hat{G})$, which will be the key to our final theorem.

Lemma A.4.2. *The following relation holds for the Fourier transform of derivatives of functions on G :*

$$(\mathcal{F}_G \circ \mathcal{A}_i f)(\sigma^{p,s}) = d\sigma^{p,s}(A_i) \mathcal{F}_G f(\sigma^{p,s}) \tag{A.28}$$

Proof. As a first step, we note the following relation for the composition of the right-regular representation and the Fourier transform:

$$\begin{aligned} (\mathcal{F}_G \circ \mathcal{R}_h f)(\sigma^{p,s}) &= \int_G f(gh) \sigma_{g^{-1}}^{p,s} dg = \int_G f(g') \sigma_{(g'h^{-1})^{-1}}^{p,s} dg' \\ &= \sigma_h^{p,s} (\mathcal{F}_G f)(\sigma^{p,s}) = \sigma_h^{p,s} (\mathcal{F}_G f)(\sigma^{p,s}). \end{aligned}$$

The desired identity then follows from linearity and continuity of the Fourier transform \mathcal{F}_G . In more detail, we have:

$$\begin{aligned} d\sigma^{p,s}(A_i) \mathcal{F}_G f(\sigma^{p,s}) &= \lim_{t \rightarrow 0} \frac{\sigma_{e^{tA_i}}^{p,s} - I}{t} \mathcal{F}_G f(\sigma^{p,s}) \\ &= \lim_{t \rightarrow 0} \frac{(\mathcal{F}_G \circ \mathcal{R}_{e^{tA_i}} f)(\sigma^{p,s}) - \mathcal{F}_G f(\sigma^{p,s})}{t} \\ &= \left(\mathcal{F}_G \circ \lim_{t \rightarrow 0} \frac{\mathcal{R}_{e^{tA_i}} f - f}{t} \right) (\sigma^{p,s}) = (\mathcal{F}_G \circ \mathcal{A}_i f)(\sigma^{p,s}), \end{aligned}$$

where subsequently we applied the definition of generator $d\sigma^{p,s}(A_i)$ (recall Definition A.4.1), we used the second identity in (C.13) for the special case $h = e^{A_i}$, we used continuity of the Fourier transform and finally identity we used the definition of $d\mathcal{R}(A_i)$. □

We conclude the appendix with the following theorem, that shows the correspondence between the Fourier transform on $SE(3)$ and the formulation used earlier in the paper based on the Fourier transform on the Cartesian product space $\mathbb{R}^3 \times S^2$.

Theorem A.4.3. *Let $j \in \{1, 2\}$ label the diffusion PDE ($j = 1$) and the convection-diffusion PDE ($j = 2$). The $SE(3)$ -Fourier transform $\hat{K}_t^j = \mathcal{F}_G \tilde{K}_t^j$ of $\tilde{K}_t^j : SE(3) \rightarrow \mathbb{R}^+$ satisfies*

$$\begin{cases} \partial_t \hat{K}_t^1(\sigma^{p,s}) = [D_{33}(d\sigma^{p,s}(A_3))^2 + D_{44}((d\sigma^{p,s}(A_4))^2 + (d\sigma^{p,s}(A_5))^2)] \hat{K}_t^1(\sigma^{p,s}), \\ \hat{K}_0^1 = 2\pi \text{id}, \end{cases} \tag{A.29}$$

whereas for $j = 2$ we have

$$\begin{cases} \partial_t \hat{K}_t^2(\sigma^{p,s}) = [-d\sigma^{p,s}(A_3) + D_{44} ((d\sigma^{p,s}(A_4))^2 + (d\sigma^{p,s}(A_5))^2)] \hat{K}_t^2(\sigma^{p,s}), \\ \hat{K}_0^2 = 2\pi \text{id}. \end{cases} \quad (\text{A.30})$$

When we set

$$p = r = \|\boldsymbol{\omega}\| \text{ and } \mathbf{u} = -\|\boldsymbol{\omega}\| \mathbf{R}_{\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}}^T \mathbf{n}, \quad (\text{A.31})$$

this system is for $j = 1$ fully equivalent to Eq. (3.82), and for $j = 2$ (up to temporal Laplace transform) fully equivalent to (3.100) with initial condition $\hat{W}(\boldsymbol{\omega}, \mathbf{n}, 0) = \delta_{\mathbf{e}_z}$. In fact the solution of the corresponding PDE-evolutions (2.22) can be rewritten as

$$\tilde{W}(\mathbf{y}, \mathbf{n}, t) = \mathcal{F}_{SE(3)}^{-1} \left[\mathcal{F}_{SE(3)} \tilde{K}_t^j \mathcal{F}_{SE(3)} \tilde{W}(\cdot, 0) \right] (\mathbf{y}, \mathbf{R}_\mathbf{n}). \quad (\text{A.32})$$

Proof. Follows by application of Lemma A.4.2 and application of the Fourier transform on $SE(3)$ to the PDE-systems for contour enhancement and contour completion. Regarding the statement of obtaining equivalent systems under our identification (A.31) we note that regarding the spatial generator that

$$\begin{aligned} (d\sigma^{p,s}(A_3)\phi)(\mathbf{u}) &= \lim_{h \downarrow 0} \frac{(\sigma_{(h\mathbf{e}_z, I)}^{p,s}\phi)(\mathbf{u}) - \phi(\mathbf{u})}{h} = -ip \left(\mathbf{e}_z \cdot \frac{\mathbf{u}}{\|\mathbf{u}\|} \right) \phi(\mathbf{u}), \\ (d\mathcal{R}(A_3)\tilde{W})(\mathbf{y}, \mathbf{R}_\mathbf{n}, t) &= \lim_{h \downarrow 0} \frac{(\mathcal{R}_{(h\mathbf{e}_z, I)}\tilde{W})(\mathbf{y}, \mathbf{R}_\mathbf{n}, t) - \tilde{W}(\mathbf{y}, \mathbf{R}_\mathbf{n}, t)}{h} = \mathcal{A}_3 \tilde{W}(\mathbf{y}, \mathbf{R}_\mathbf{n}, t), \\ \mathcal{F}_{\mathbb{R}^3} [d\mathcal{R}(A_3)\tilde{W}(\cdot, \mathbf{R}_\mathbf{n}, t)](\boldsymbol{\omega}) &= (\mathcal{F}_{\mathbb{R}^3} \mathbf{n} \cdot \nabla W(\cdot, \mathbf{n}, t))(\boldsymbol{\omega}) = ir \left(\mathbf{n} \cdot \frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|} \right) [\mathcal{F}_{\mathbb{R}^3} W(\cdot, \mathbf{n}, t)](\boldsymbol{\omega}). \end{aligned} \quad (\text{A.33})$$

So in order to match the multiplier operators at the right hand side of (A.33) we indeed need (A.31), where we note that $\mathbf{R}_{\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}} \mathbf{e}_z = \frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}$.

Regarding the angular generators we note that the following generator is independent of s :

$$\begin{aligned} (d\sigma^{p,s}(A_4))^2 + (d\sigma^{p,s}(A_5))^2 &= \mathcal{F}_G \circ ((d\mathcal{R}(A_4))^2 + (d\mathcal{R}(A_5))^2) \circ \mathcal{F}_G^{-1} \\ &= \mathcal{F}_G \circ ((\mathcal{A}_4)^2 + (\mathcal{A}_5)^2) \circ \mathcal{F}_G^{-1}, \end{aligned}$$

and recall that

$$\left((\mathcal{A}_4^2 + \mathcal{A}_5^2) \tilde{W} \right) (\mathbf{y}, \mathbf{R}_\mathbf{n}, t) = \left((\mathcal{A}_4^2 + \mathcal{A}_5^2 + \mathcal{A}_6^2) \tilde{W} \right) (\mathbf{y}, \mathbf{R}_\mathbf{n}, t) = (\Delta_{S^2} W)(\mathbf{y}, \mathbf{n}, t).$$

As a result, the generators of the linear evolutions relate via conjugation with the Fourier transform on $SE(3)$

and thereby the same applies to the evolutions themselves. Finally, regarding (A.32), the expression is invariant under the choice of rotation $\mathbf{R}_\mathbf{n}$ mapping \mathbf{e}_z onto \mathbf{n} (since only

the matrix coefficients with $m = m' = 0$ contribute to the series (A.24) of the inverse Fourier transformation).

□

Remark A.4.4. *Akin to the $SE(2)$ -case (where the substitution showing the equivalence was similar, cf. [DvA08, App.A, Thm A.2]) we conclude that the tool of Fourier transform on $SE(3)$ is not strictly needed to find an explicit analytic description of the heat-kernels. Nevertheless, it is clearly present in our specific choice of coordinates (3.25) in Lemma 3.2.2 and Fig. 3.1, that was crucial in our derivation of the exact solutions. In fact, application of the Fourier transform $\mathcal{F}_{SE(3)}$ on the Lie group $SE(3)$ yields isomorphic equations to the ones obtained by application of the operator*

$$(\text{id}_{\mathbb{L}_2(\mathbb{R}^3)} \otimes \mathcal{F}_{S^2}) \circ \Xi \circ (\mathcal{F}_{\mathbb{R}^3} \otimes \text{id}_{\mathbb{L}_2(S^2)}) ,$$

with operator $\Xi : \mathbb{L}_2(\mathbb{R}^3 \times S^2) \rightarrow \mathbb{L}_2(\mathbb{R}^3 \times S^2)$ given by

$$(\Xi \hat{U})(\boldsymbol{\omega}, \mathbf{n}) = \hat{U}(\boldsymbol{\omega}, \mathbf{R}_{\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}}^T \mathbf{n}),$$

which relates to our substitution (A.31), and which is the actual net operator we applied in our classical analytical approach pursued in the body of the article (Lemma 3.2.2 and Figure 3.1). We conclude that in retrospect, in our classical analytic PDE-approach, we did solve the PDE's via Fourier transform on $SE(3)$.

B

Appendix to Chapter 5

B.1 Well-posedness and convergence of the Reeds-Shepp models

We introduce in §B.1.1 some general elements of control theory, which are specialized in §B.1.2 to the Reeds-Shepp models and their approximations.

B.1.1 Closedness of controllable paths

In this section, we introduce the notion of an admissible path γ with respect to some controls \mathfrak{B} . We state in Theorem B.1.6 a closedness result, slightly generalizing the one from [Che16], from which we deduce in Corollaries B.1.8 and B.1.9 an existence and a convergence result for a minimum time optimal control problem. The first ingredient of this approach is the notion of Hausdorff distance on a metric space.

Definition B.1.1. *Given a metric space \mathbb{E} , we let $\mathcal{K}(\mathbb{E})$ be the collection of non-empty compact subsets of \mathbb{E} . The distance function $d_A : \mathbb{E} \rightarrow \mathbf{R}_+$ and the Hausdorff distance $\mathcal{H}(A, B)$, where $A, B \in \mathcal{K}(\mathbb{E})$, are defined respectively by*

$$d_A(x) := \inf_{y \in A} d(x, y), \quad \mathcal{H}(A, B) := \max\{\sup_{x \in B} d_A(x), \sup_{x \in A} d_B(x)\}.$$

In the following, we fix a closed set \mathbf{X} , contained in an Euclidean vector space \mathbb{E} , or in a complete Riemannian manifold \mathbb{M} . In the applications considered in this paper, \mathbf{X} is of the form $\mathbf{X}_0 \times S^{d-1}$, where $\mathbf{X}_0 \subset \mathbb{R}^d$ is some image domain, see Fig. 7.4, or the set of accessible points in a map (which excludes the walls), see Fig. 7.3. The embedding space can be the vector space $\mathbb{E} = \mathbb{R}^d \times \mathbb{R}^d$, which is an acceptable but rather extrinsic point of view, or the Riemannian manifold $\mathbb{M} = \mathbb{R}^d \times S^{d-1}$, equipped with the metric \mathcal{G}_ε for some arbitrary but fixed $\varepsilon > 0$, see (5.24).

We equip the collection of all Lipschitz paths $\Gamma := \text{Lip}([0, 1], \mathbf{X})$ with the topology of uniform convergence. We will make use of Ascoli's lemma [Asc83, Arz96], which states that any uniformly bounded and equicontinuous sequence of paths admits a converging sub-sequence. In our case the paths are Lipschitz with a common Lipschitz constant.

Definition B.1.2. *Given a normed vector space V , we denote by $\mathfrak{C}(V) \subset \mathcal{K}(V)$ the collection of non-empty compact subsets of V , which are convex and contained in the unit ball.*

Remark B.1.3. *The restriction to convex subsets is essential. For a uniformly converging sequence of Lipschitz functions $\gamma_n : [0, 1] \rightarrow \mathbb{M}$ with limit γ_* , with $\dot{\gamma}_n(t) \in K$ for a.e. $t \in [0, 1]$ and K a compact set, we can deduce that $\dot{\gamma}_* \in \text{Hull}(K)$, for a.e. $t \in [0, 1]$. The convexity then guarantees that $\dot{\gamma}_* \in K = \text{Hull}(K)$.*

Definition B.1.4. A family of controls \mathcal{B} on the set \mathbf{X} is an element of the set \mathfrak{B} defined by

- If $\mathbf{X} \subset \mathbb{E}$ an Euclidean vector space, then $\mathfrak{B} := C^0(\mathbf{X}, \mathfrak{C}(\mathbb{E}))$.
- If $\mathbf{X} \subset \mathbb{M}$ a Riemannian manifold, then $\mathfrak{B} := \{\mathcal{B} \in C^0(\mathbf{X}, \mathcal{K}(T\mathbb{M})) \mid \forall \mathbf{p} \in \mathbf{X}, \mathcal{B}(\mathbf{p}) \in \mathfrak{C}(T_{\mathbf{p}}\mathbb{M})\}$.

In both cases, \mathfrak{B} is equipped with the topology of locally uniform convergence.

Definition B.1.5. A path γ is $T\mathcal{B}$ -admissible, where $\gamma \in \Gamma$, $T \in \mathbf{R}_+$ and $\mathcal{B} \in \mathfrak{B}$, iff for almost every $t \in [0, 1]$

$$\dot{\gamma}(t) \in T\mathcal{B}(\gamma(t)).$$

We denoted $T\mathcal{B} := \{T\mathbf{v} \mid \mathbf{v} \in B\}$, where $T \in \mathbf{R}_+$ and B is a subset of a vector space. Note the potential conflict of notation with the tangent space $T\mathbb{M}$ to the embedding manifold \mathbb{M} , which should be clear from context. If a path γ is $T\mathcal{B}$ -admissible for some controls $\mathcal{B} \in \mathfrak{B}$, then it must be T -Lipschitz. The following result slightly extends, for our convenience, Corollary A.5 in [Che16].

Theorem B.1.6. The set $\{(\gamma, T, \mathcal{B}) \in \Gamma \times \mathbf{R}_+ \times \mathfrak{B} \mid \gamma \text{ is } T\mathcal{B}\text{-admissible}\}$ is closed.

Proof. Let $(\gamma_n, T_n, \mathcal{B}_n)$ be sequences of paths, times and controls converging to $(\gamma_\infty, T_\infty, \mathcal{B}_\infty)$, and such that γ_n is $T_n\mathcal{B}_n$ -admissible for all $n \geq 0$. Since the paths γ_n are converging as $n \rightarrow \infty$, they lay in a common compact subset \mathbf{X}' of the closed domain \mathbf{X} , recall Remark B.1.3. As a result, the restricted controls $\mathcal{B}'_n := (\mathcal{B}_n|_{\mathbf{X}'})$ are uniformly converging as $n \rightarrow \infty$. In the case where $\mathbf{X} \subset \mathbb{E}$ a Euclidean space, applying Corollary A.5 in [Che16] to the sequence $(\gamma_n, T_n\mathcal{B}'_n)$ we obtain that γ_∞ is $T_\infty\mathcal{B}_\infty$ -admissible as announced.

In the case where $\mathbf{X} \subset \mathbb{M}$ a Riemannian manifold, an additional proof ingredient is required. Let \mathbb{M}' be an open neighborhood of \mathbf{X}' with compact closure in \mathbb{M} , and let $\mathcal{I} : \mathbb{M}' \rightarrow \mathbb{E}$ be an embedding (i.e. an injective immersion) with bounded distortion of the manifold \mathbb{M}' into a Euclidean space \mathbb{E} of sufficiently high dimension, which by Whitney's embedding theorem is known to exist. Define the set $\mathbf{X}'' := \mathcal{I}(\mathbf{X}')$, the paths $\gamma''_n := \mathcal{I} \circ \gamma_n$, and controls $\mathcal{B}''_n(\mathcal{I}(\mathbf{p})) := d\mathcal{I}(\mathbf{p}, \mathcal{B}_n(\mathbf{p}))$ for all $\mathbf{p} \in \mathbf{X}'$ and $n \in \mathbf{N} \cup \{\infty\}$. Applying again Corollary A.5 in [Che16] we obtain that γ''_∞ is $T_\infty\mathcal{B}''_\infty$ admissible, hence that γ_∞ is $T_\infty\mathcal{B}_\infty$ -admissible as announced. \square

In line with the identity (5.9), we rely on the following definition where we rescale the time interval to $[0, 1]$.

Definition B.1.7. For any $\mathcal{B} \in \mathfrak{B}$, $\mathbf{p}, \mathbf{q} \in \mathbf{X}$, we let

$$T_{\mathcal{B}}(\mathbf{p}, \mathbf{q}) := \inf\{T \geq 0 \mid \exists \gamma \in \Gamma, \gamma(0) = \mathbf{p}, \gamma(1) = \mathbf{q}, \text{ and } \gamma \text{ is } T\mathcal{B}\text{-admissible}\}. \quad (\text{B.1})$$

Corollary B.1.8. If $\mathcal{B} \in \mathfrak{B}$, $\mathbf{p}, \mathbf{q} \in \mathbf{X}$ are such that $T_{\mathcal{B}}(\mathbf{p}, \mathbf{q}) < \infty$, then the inf. (B.1) is attained.

Proof. Let $T := T_{\mathcal{B}}(\mathbf{p}, \mathbf{q})$, and for each $0 < \varepsilon \leq 1$ let γ_{ε} be a $(T + \varepsilon)\mathcal{B}$ -admissible path from \mathbf{p} to \mathbf{q} , which is thus $(T + 1)$ -Lipschitz. By Arzela-Ascoli's lemma [Asc83, Arz96] there exists a converging sequence of paths $\gamma_{\varepsilon_n} \rightarrow \gamma_0$ as $n \rightarrow \infty$. The limit path γ_0 is $T\mathcal{B}$ -admissible by Theorem B.1.6, and the result follows. \square

Corollary B.1.9. For all $\varepsilon \in [0, 1]$ let $\mathcal{B}_{\varepsilon} \in \mathfrak{B}$. Assume that $\mathcal{B}_{\varepsilon} \rightarrow \mathcal{B}_0$ as $\varepsilon \rightarrow 0$, and that $\mathcal{B}_{\varepsilon}(\mathbf{p}) \supset \mathcal{B}_0(\mathbf{p})$ for all $\varepsilon \geq 0$, $\mathbf{p} \in \mathbf{X}$. Then

$$T_{\mathcal{B}_{\varepsilon}}(\mathbf{p}, \mathbf{q}) \rightarrow T_{\mathcal{B}_0}(\mathbf{p}, \mathbf{q}), \quad \text{as } \varepsilon \rightarrow 0.$$

Let $T_{\varepsilon} := T_{\mathcal{B}_{\varepsilon}}(\mathbf{p}, \mathbf{q})$ for each $\varepsilon \geq 0$. Assume in addition that there exists a unique $T_0\mathcal{B}_0$ -admissible path γ_0 from \mathbf{p} to \mathbf{q} , and for each $\varepsilon > 0$ denote by γ_{ε} an arbitrary path from \mathbf{p} to \mathbf{q} which is $(\varepsilon + T_{\varepsilon})\mathcal{B}_{\varepsilon}$ admissible. Then $\gamma_{\varepsilon} \rightarrow \gamma_0$ as $\varepsilon \rightarrow 0$.

Proof. The inclusion $\mathcal{B}_{\varepsilon}(\mathbf{p}) \subset \mathcal{B}_0(\mathbf{p})$, $\forall \mathbf{p} \in \mathbb{M}$, implies the inequality $T_{\varepsilon} \leq T_0$, for all $\varepsilon \geq 0$. Denoting $T_* := \limsup T_{\varepsilon}$ as $\varepsilon \rightarrow 0$, we thus observe that $T_* \leq T_0$. For the reverse inequality $T_* \geq T_0$, we apply Arzela-Ascoli lemma to the family of paths $(\gamma_{\varepsilon})_{0 < \varepsilon \leq 1}$ which are $(T_0 + 1)$ -Lipschitz by construction, and obtain a converging subsequence of paths $\gamma_{\varepsilon_n} \rightarrow \gamma_*$. Theorem B.1.6 implies the admissibility of γ_* with respect to the controls $T_*\mathcal{B}_0$. Thus $T_* \geq T_0$ but since $T_* \leq T_0$, we must have $T_* = T_0$, and $\gamma_* = \gamma_0$ by the uniqueness assumption. The result follows. \square

More generally, if the infimum (B.1) is realized by a family $(\gamma_i)_{i \in I}$ of paths, then for any sequence $\varepsilon_n \rightarrow 0$ one can find a subsequence such that $\gamma_{\varepsilon_{\varphi(n)}} \rightarrow \gamma_i$ as $n \rightarrow \infty$ for some $i \in I$.

B.1.2 Specialization to the Reeds-Shepp models

We begin this section by recalling, and slightly generalizing, the notion of Finsler metric introduced in §5.2.2. We then prove that the Reeds-Shepp metrics \mathcal{F}_0 and \mathcal{F}_0^+ are indeed Finsler metrics in this sense.

Definition B.1.10. A metric on a complete Riemannian manifold \mathbb{M} is a map $\mathcal{F} : T\mathbb{M} \rightarrow [0, +\infty]$. With respect to the second variable, it must be 1-homogeneous, convex, and bounded below by $\delta\|\cdot\|$, where δ is a positive constant. In terms of regularity, the sets

$\mathcal{B}_{\mathcal{F}}(\mathbf{p}) := \{\dot{\mathbf{p}} \in T_{\mathbf{p}}\mathbb{M} \mid \mathcal{F}(\mathbf{p}, \dot{\mathbf{p}}) \leq 1\}$ must be closed and depend continuously on $\mathbf{p} \in \mathbb{M}$ with respect to the Hausdorff distance on $T\mathbb{M}$.

The next proposition is due to (5.9).

Proposition B.1.11. *With the notations of Definition B.1.10, the sets $\mathbf{p} \in \mathbb{M} \mapsto \mathcal{B}_{\mathcal{F}}(\mathbf{p})$ form a family of controls on $(\mathbb{M}, \delta\|\cdot\|)$. In addition for all $\mathbf{p}, \mathbf{q} \in \mathbb{M}$*

$$d_{\mathcal{F}}(\mathbf{p}, \mathbf{q}) = T_{\mathcal{B}_{\mathcal{F}}}(\mathbf{p}, \mathbf{q}).$$

Proposition B.1.12. *The Reeds-Shepp metrics $(\mathcal{F}_{\varepsilon})_{0 \leq \varepsilon \leq 1}$ and $(\mathcal{F}_{\varepsilon}^+)_{0 \leq \varepsilon \leq 1}$ are indeed metrics in the sense of Definition B.1.10, for any $\varepsilon \in [0, 1]$. The associated controls $\mathcal{B}_{\varepsilon} := \mathcal{B}_{\mathcal{F}_{\varepsilon}}, \mathcal{B}_{\varepsilon}^+ := \mathcal{B}_{\mathcal{F}_{\varepsilon}^+}$ depend continuously on the parameter $\varepsilon \in [0, 1]$, and satisfy the inclusions $\mathcal{B}_{\varepsilon}(\mathbf{p}) \subset \mathcal{B}_{\varepsilon'}(\mathbf{p})$ and $\mathcal{B}_{\varepsilon}^+(\mathbf{p}) \subset \mathcal{B}_{\varepsilon'}^+(\mathbf{p})$ for any $\mathbf{p} \in \mathbb{M}$ and $0 \leq \varepsilon \leq \varepsilon' \leq 1$.*

Proposition B.1.12 allows to apply the results of §B.1.1 to the Reeds-Shepp metrics. Theorem 5.2.3 then directly follows from Corollary B.1.9. The only remaining non-trivial claim in Proposition B.1.12 is the continuity of the controls on \mathbb{M} , recall Definitions B.1.4, and their convergence $\mathcal{B}_{\varepsilon} \rightarrow \mathcal{B}_0$ as $\varepsilon \rightarrow 0$, as required in Corollary B.1.9. These two properties are implied by the continuity on $[0, 1] \times \mathbb{M}$, that we next prove, of the following maps

$$\begin{aligned} [0, 1] \times \mathbb{M} \ni (\varepsilon, \mathbf{p}) &\rightarrow \mathcal{B}_{\varepsilon}(\mathbf{p}) \in \mathfrak{C}(T_{\mathbf{p}}\mathbb{M}), \\ [0, 1] \times \mathbb{M} \ni (\varepsilon, \mathbf{p}) &\rightarrow \mathcal{B}_{\varepsilon}^+(\mathbf{p}) \in \mathfrak{C}(T_{\mathbf{p}}\mathbb{M}), \end{aligned} \tag{B.2}$$

with $\mathfrak{C}(T_{\mathbf{p}}\mathbb{M})$ defined in Definition B.1.2 and equipped with the Hausdorff distance.

Lemma B.1.13. *Let B be a compact subset of a metric space \mathbb{E} , and let $\varphi \in C^0(B, \mathbb{E})$. Then*

$$\mathcal{H}(B, \varphi(B)) \leq \sup_{x \in B} d(x, \varphi(x)).$$

This basic lemma, stated without proof, is used in the next lemma to obtain an explicit estimate of the Hausdorff distance between the controls sets of the Reeds-Shepp models.

Lemma B.1.14. *Let $\mathbf{n}_1, \mathbf{n}_2 \in S^{d-1}$, let $a_1, a_2, b_1, b_2 \geq 1$, and let $\varepsilon_1, \varepsilon_2 \in [0, 1]$. For each $i \in \{1, 2\}$, let B_i be the collection of all $(\dot{\mathbf{x}}, \dot{\mathbf{n}}) \in \mathbb{R}^d \times \mathbb{R}^d$ obeying*

$$\begin{cases} \dot{\mathbf{n}} \cdot \mathbf{n}_i = 0, \\ \left\{ \begin{array}{ll} a_i^2 \|\dot{\mathbf{n}}\|^2 + b_i^2 (|\dot{\mathbf{x}} \cdot \mathbf{n}_i|^2 + \varepsilon_i^{-2} \|\dot{\mathbf{x}} \wedge \mathbf{n}_i\|^2) \leq 1, & \varepsilon_i > 0 \\ a_i^2 \|\dot{\mathbf{n}}\|^2 + b_i^2 |\dot{\mathbf{x}} \cdot \mathbf{n}_i|^2 \leq 1 \text{ and } \dot{\mathbf{x}} \wedge \mathbf{n}_i = 0, & \varepsilon_i = 0. \end{array} \right. \end{cases}$$

$$\text{Then } \mathcal{H}(B_1, B_2) \leq |a_1^{-1} - a_2^{-1}| + |b_1^{-1} - b_2^{-1}| + \sqrt{2(1 - \mathbf{n}_1 \cdot \mathbf{n}_2)} + |\varepsilon_1 - \varepsilon_2|. \tag{B.3}$$

The same estimate holds for the sets B_i^+ , $i \in \{1, 2\}$, defined by the inequalities

$$\begin{cases} \mathbf{\dot{n}} \cdot \mathbf{n}_i = 0, \\ a_i^2 \|\mathbf{\dot{n}}\|^2 + b_i^2 ((\mathbf{\dot{x}} \cdot \mathbf{n}_i)_+^2 + \varepsilon_i^{-2} (\|\mathbf{\dot{x}} \wedge \mathbf{n}_i\|^2 + (\mathbf{\dot{x}} \cdot \mathbf{n}_i)_-^2)) \leq 1, \text{ if } \varepsilon_i > 0, \\ a_i^2 \|\mathbf{\dot{n}}\|^2 + b_i^2 (\mathbf{\dot{x}} \cdot \mathbf{n}_i)_+^2 \leq 1 \text{ and } \mathbf{\dot{x}} \wedge \mathbf{n}_i = 0, \quad \mathbf{\dot{x}} \cdot \mathbf{n}_i \geq 0, \text{ if } \varepsilon_i = 0. \end{cases}$$

Proof. It suffices to establish the announced estimate (B.3) when the tuples $(a_i, b_i, \mathbf{n}_i, \varepsilon_i)$, $i \in \{1, 2\}$, differ by a *single* element of the four, and then to use the subadditivity of the Hausdorff distance. In each case we apply Lemma B.1.13 to a well chosen surjective map $\varphi : B_1 \rightarrow B_2$ (resp $\varphi^+ : B_1^+ \rightarrow B_2^+$).

- Case $a_1 \neq a_2$. Assume w.l.o.g. that $a_1 < \infty$, and observe that for all $(\mathbf{\dot{x}}, \mathbf{\dot{n}}) \in B_1$ one has $a_1 \|\mathbf{\dot{x}}\| \leq 1$, hence $\|a_1 \mathbf{\dot{x}}/a_2 - \mathbf{\dot{x}}\| \leq |a_1^{-1} - a_2^{-1}|$. Choose $\varphi(\mathbf{\dot{x}}, \mathbf{\dot{n}}) := (a_1 \mathbf{\dot{x}}/a_2, \mathbf{\dot{n}})$.
- Case $b_1 \neq b_2$. As above, with $\varphi(\mathbf{\dot{x}}, \mathbf{\dot{n}}) := (\mathbf{\dot{x}}, b_1 \mathbf{\dot{n}}/b_2)$, yielding the upper bound $|b_1^{-1} - b_2^{-1}|$.
- Case $\mathbf{n}_1 \neq \mathbf{n}_2$. Let R be the rotation of \mathbb{R}^d which maps \mathbf{n}_1 onto \mathbf{n}_2 , in such a way that it maps the space orthogonal to the plane $\text{Span}(\mathbf{n}_1, \mathbf{n}_2)$ onto itself. A simple calculation yields $\|R - \text{Id}\| = 2 \sin[\frac{1}{2} \cos^{-1}(\mathbf{n}_1 \cdot \mathbf{n}_2)] = \sqrt{2(1 - \mathbf{n}_1 \cdot \mathbf{n}_2)}$. The result follows by choosing $\varphi(\mathbf{\dot{x}}, \mathbf{\dot{n}}) := (R\mathbf{\dot{x}}, R\mathbf{\dot{n}})$, so that $\|\varphi(\mathbf{\dot{x}}, \mathbf{\dot{n}}) - (\mathbf{\dot{x}}, \mathbf{\dot{n}})\| \leq \|R - \text{Id}\| \sqrt{\|\mathbf{\dot{n}}\|^2 + \|\mathbf{\dot{x}}\|^2} \leq \|R - \text{Id}\|$ for all $(\mathbf{\dot{x}}, \mathbf{\dot{n}}) \in B_1$ as announced.
- Case $\varepsilon_1 \neq \varepsilon_2$. Assume w.l.o.g. that $\varepsilon_1 > 0$, and consider the orthogonal projections

$$P_1(\mathbf{\dot{x}}) := (\mathbf{\dot{x}} \cdot \mathbf{n}_1) \mathbf{n}_1 \qquad P_1^\perp(\mathbf{\dot{x}}) := (\text{Id} - P_1)(\mathbf{\dot{x}}).$$

Note that $P_1^\perp(\mathbf{\dot{x}}) \leq \varepsilon_1$ if $(\mathbf{\dot{x}}, \mathbf{\dot{n}}) \in B_1$, and that $\|\mathbf{\dot{x}}\| \leq \varepsilon_1$ if $(\mathbf{\dot{x}}, \mathbf{\dot{n}}) \in B_1^+$ and $\mathbf{\dot{x}} \cdot \mathbf{n}_1 \leq 0$. The result follows by choosing

$$\begin{aligned} \varphi(\mathbf{\dot{x}}, \mathbf{\dot{n}}) &:= \left(P_1(\mathbf{\dot{x}}) + \frac{\varepsilon_2}{\varepsilon_1} P_1^\perp(\mathbf{\dot{x}}), \mathbf{\dot{n}} \right), \\ \varphi^+(\mathbf{\dot{x}}, \mathbf{\dot{n}}) &:= \begin{cases} \varphi(\mathbf{\dot{x}}, \mathbf{\dot{n}}) & \text{if } \mathbf{\dot{x}} \cdot \mathbf{n}_1 \geq 0, \\ \left(\frac{\varepsilon_2}{\varepsilon_1} \mathbf{\dot{x}}, \mathbf{\dot{n}} \right) & \text{otherwise.} \end{cases} \quad \square \end{aligned}$$

Proof of Proposition B.1.12. Since working with Hausdorff distances on the abstract tangent bundle $T\mathbb{M}$ is not very practical, we make use of the canonical embedding $\mathcal{I} : \mathbb{R}^d \times S^{d-1} \rightarrow \mathbb{R}^d \times \mathbb{R}^d$ of the manifold \mathbb{M} into the Euclidean vector space \mathbb{R}^{2d} given by $(\mathbf{x}, \mathbf{n}) \mapsto (\mathbf{x}, \mathbf{n})$, which has bounded distortion. It suffices to prove the continuity of the image of the control sets $(\varepsilon, \mathbf{p}) \rightarrow d\mathcal{I}(\mathbf{p}, \mathcal{B}_{\mathcal{F}_\varepsilon}(\mathbf{p}))$ (resp. likewise with $\mathcal{F}_\varepsilon^+$) by the tangent maps to this embedding, which follows by Lemma B.1.14. Indeed the lemma shows that

$$((\varepsilon_1, \mathbf{p}_1) \rightarrow (\varepsilon_2, \mathbf{p}_2)) \implies (\mathcal{H}(B_{\mathcal{F}_{\varepsilon_1}}, B_{\mathcal{F}_{\varepsilon_2}}) \rightarrow 0),$$

and it includes the spherical constraint via the velocity constraint $\dot{\mathbf{n}} \cdot \mathbf{n}_i = \frac{d}{dt}(\mathbf{n}(t) \cdot \mathbf{n}(t))|_{t=0} = 0$ for a smooth curve $\gamma(t) = (\mathbf{x}(t), \mathbf{n}(t))$ passing through $\gamma(0) = (\mathbf{x}_i, \mathbf{n}_i)$. \square

B.2 Backtracking of geodesics in $(\mathbb{M}, d_{\mathcal{F}})$

This section is devoted to a generic ingredient in the proof of Theorem 5.2.15, regarding backtracking of Geodesics in the (quasi)-Metric Space $(\mathbb{M}, d_{\mathcal{F}})$ in general. Although, these results are standard in Finsler Geometry, we aim to provide a concise overview.

Lemma B.2.1. *Let F be an asymmetric norm on a vector space \mathbb{E} , and assume that F^* is differentiable at $\hat{\mathbf{p}} \in \mathbb{E}^*$. Then*

$$F(dF^*(\hat{\mathbf{p}})) = 1, \quad \langle \hat{\mathbf{p}}, dF^*(\hat{\mathbf{p}}) \rangle = F^*(\hat{\mathbf{p}}).$$

Proof. The 1st claim follows by differentiation of F^*

$$F^*(\hat{\mathbf{p}}) = \sup_{\dot{\mathbf{p}} \in \mathbb{E} \setminus \{0\}} \frac{\langle \hat{\mathbf{p}}, \dot{\mathbf{p}} \rangle}{F(\dot{\mathbf{p}})} = \max_{F(\dot{\mathbf{p}})=1} \langle \hat{\mathbf{p}}, \dot{\mathbf{p}} \rangle.$$

The 2nd claim is Euler's formula for homogeneous functions. \square

Proposition B.2.2. *Let $\mathbf{p}_S, \mathbf{p}_T \in \mathbb{M}$, let γ be a minimizing geodesic from \mathbf{p}_S to \mathbf{p}_T w.r.t. a continuous metric \mathcal{F} , and let $t \in [0, 1]$. Assume that the distance map U from \mathbf{p}_S is differentiable at $\gamma(t)$, and that the dual metric \mathcal{F}^* is differentiable w.r.t. the second variable at $(\gamma(t), dU(\gamma(t)))$. Then γ is differentiable at time t and with $L := d_{\mathcal{F}}(\mathbf{p}_S, \mathbf{p}_T)$*

$$\dot{\gamma}(t) = L \, d_{\hat{\mathbf{p}}} \mathcal{F}^*(\gamma(t), dU(\gamma(t))), \quad \gamma(0) = \mathbf{p}_S, \gamma(1) = \mathbf{p}_T. \quad (\text{B.4})$$

Proof. The path γ has constant speed L , and $t \mapsto U(\gamma(t))$ increases linearly from 0 to L on it. Let $t \in [0, 1]$ be as in the statement of the proposition, and let

$$\dot{\Gamma}(t) := \lim_{n \rightarrow \infty} (\gamma(t + \varepsilon_n) - \gamma(t)) / \varepsilon_n$$

for some sequence $\varepsilon_n \rightarrow 0$. Then

$$\mathcal{F}(\gamma(t), \dot{\Gamma}(t)) = L \quad \text{and} \quad \langle dU(\gamma(t)), \dot{\Gamma}(t) \rangle = L.$$

For typographic simplicity let us denote $\mathbf{p} := \gamma(t)$, $\dot{\mathbf{p}} := \dot{\Gamma}(t)$, $F = \mathcal{F}(\mathbf{p}, \cdot)$ and $F^* := \mathcal{F}^*(\mathbf{p}, \cdot)$. By Lemma B.2.1 and the eikonal equation (5.5), the vector $\dot{\mathbf{q}} = dF^*(dU(\mathbf{p}))$ obeys

$$F(\dot{\mathbf{q}}) = F(dF^*(dU(\mathbf{p}))) = 1, \\ \langle dU(\mathbf{p}), \dot{\mathbf{q}} \rangle = \langle dU(\mathbf{p}), dF^*(dU(\mathbf{p})) \rangle = F^*(dU(\mathbf{p})) = 1.$$

Note that the duality-bracket/norm inequality is saturated by

$$\langle dU(\mathbf{p}), \dot{\mathbf{q}} \rangle = 1 = F^*(dU(\mathbf{p}))F(\dot{\mathbf{q}}),$$

and that the assumed differentiability of the dual norm F^* at the point $\hat{\mathbf{p}} = dU(\mathbf{p})$ implies the strict convexity of the primal norm F (up to 1-homogeneity) at the point $dF^*(\hat{\mathbf{p}}) = \dot{\mathbf{q}}$. Hence $\dot{\mathbf{q}}$ is the unique solution to the system “ $F^*(\dot{\mathbf{q}}) = 1$ and $\langle dU(\mathbf{p}), \dot{\mathbf{q}} \rangle = 1$ ”, and therefore $\dot{\Gamma} = L\dot{\mathbf{q}}$. This implies the differentiability of γ at time t , and the announced equality (B.4). \square

Remark B.2.3 (Lagrangians and Hamiltonians). *Given an arbitrary Finsler metric \mathcal{F} on \mathbb{M} , its half-square $\mathfrak{L} := \frac{1}{2}\mathcal{F}^2 : T(\mathbb{M}) \rightarrow [0, +\infty]$ is usually called the Lagrangian. The shortest path problem (5.1) can be reformulated in terms of the Lagrangian, thanks to the Cauchy-Schwartz’s inequality which gives*

$$d_{\mathcal{F}}(\mathbf{p}, \mathbf{q})^2 = \inf \left\{ \int_0^1 \mathcal{F}(\gamma(t), \dot{\gamma}(t))^2 dt \mid \gamma \in \text{Lip}([0, 1], \mathbb{M}), \gamma(0) = \mathbf{p}, \gamma(1) = \mathbf{q} \right\}. \quad (\text{B.5})$$

A path γ is a minimizer of (B.5) iff it is simultaneously normalized and a minimizer of (5.1). The Hamiltonian \mathfrak{H} is the Legendre-Fenchel transform of its Lagrangian \mathfrak{L} w.r.t. the second variable, hence $\mathfrak{H} = \frac{1}{2}(\mathcal{F}^*)^2$ (for details see [BCS00, ch.14.8]) The eikonal equation can thus be rephrased in terms of the Hamiltonian:

$$\mathcal{F}^*(\mathbf{p}, dU(\mathbf{p})) = 1 \Leftrightarrow \mathfrak{H}(\mathbf{p}, dU(\mathbf{p})) = \frac{1}{2}.$$

The Hamiltonian can also be used to reformulate the backtracking ODE of geodesics, thanks to the following identity which follows from the eikonal equation: for any $\mathbf{p} \in \mathbb{M}$

$$d_{\hat{\mathbf{p}}}\mathfrak{H}(\mathbf{p}, dU(\mathbf{p})) = \mathcal{F}^*(\mathbf{p}, dU(\mathbf{p})) d_{\hat{\mathbf{p}}}\mathcal{F}^*(\mathbf{p}, dU(\mathbf{p})) = d_{\hat{\mathbf{p}}}\mathcal{F}^*(\mathbf{p}, dU(\mathbf{p})). \quad (\text{B.6})$$

In geometric control theory this Hamiltonian is often referred to the ‘fixed time Hamiltonian of the action functional’, cf. [AS04, BDMS15, Sac11], and is typically used [MS10] in the Pontryagin maximum principle [AS04] for (sub-)Riemannian geodesics.

B.3 Characterization of cusps: Proof of Lemma 5.2.6

Consider Lemma 5.2.6. The structure of this lemma is $a \iff b \iff c$. The implication $a \Rightarrow b$ is trivial. The equivalence $b \Leftrightarrow c$ follows by Theorems 5.2.15, 5.2.3. The implication $b \Rightarrow a$ remains.

Suppose the d -th spatial control aligned with $\mathbf{n}(t_0)$, recall (5.18), vanishes: $\tilde{u}(t_0) = 0$. Now we show by contradiction that in this case $\dot{\tilde{u}}(t_0) \neq 0$. Suppose $\tilde{u}(t_0) = \dot{\tilde{u}}(t_0) = 0$. Then by application of Pontryagin’s Maximum Principle, similar to [BDMS15, App.A],

[DGDHS14]) and coercivity/invertibility of the SR-metric tensor $\mathcal{G}_0|_{\gamma(t_0)}$, recall (5.24), constrained to the horizontal part of the tangent space $\Delta|_{\gamma(t)} = \{(\mathbf{p}_0 = (\mathbf{x}_0, \mathbf{n}_0), \dot{\mathbf{p}}_0 = (\dot{\mathbf{x}}_0, \dot{\mathbf{n}}_0)) \in T(\mathbb{M}) \mid \mathbf{n}_0 \equiv \dot{\mathbf{x}}_0\}$, that the (analytic) spatial control variable $\tilde{u} = \mathcal{C}_1^{-2}\tilde{\lambda}$ vanishes for all times (for $d = 2$ this is directly deduced from the pendulum phase portrait [MS10] in momentum space). This leaves only purely angular momentum and motion, contradicting $\dot{\mathbf{x}}(\cdot) \neq \mathbf{0}$ in Lemma 5.2.6.

Next we verify $\tilde{u}(t_0) = \dot{\tilde{u}}(t_0) = 0 \Rightarrow \dot{\tilde{\lambda}}(t_0) = 0 = \tilde{\lambda}(t_0)$. By the chain rule for differentiation (applied to the d -th spatial momentum component $\tilde{\lambda}(t) = \langle \lambda(t), (\mathbf{n}(t), \mathbf{0}) \rangle$):

$$\begin{aligned} \left. \frac{d}{dt} \tilde{\lambda}(t) \right|_{t=t_0} &= \left. \frac{d}{dt} (\mathcal{C}_1(\gamma(t)))^{-2} \tilde{u}(t) \right|_{t=t_0} \\ &= \left. \frac{d}{dt} (\mathcal{C}_1(\gamma(t)))^{-2} \right|_{t=t_0} \tilde{u}(t_0) + \left. \frac{d}{dt} (\mathcal{C}_1(\gamma(t)))^{-2} \right|_{t=t_0} \dot{\tilde{u}}(t_0) = 0. \end{aligned}$$

We deduce from PMP's Hamiltonian equations (cf. [DGDHS14]) that

$$\dot{\tilde{\lambda}}(t_0) = \tilde{\lambda}(t_0) = 0 \Rightarrow \tilde{\lambda}(\cdot) = 0 \Rightarrow \tilde{u}(\cdot) = 0.$$

Bibliography

- [AB05] Assaf, Y. and Basser, P. J., *Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain*, NeuroImage **27** (2005), no. 1, 48–58.
- [ABB16] Agrachev, A. A., Barilari, D., and Boscain, U., *Introduction to Riemannian and Sub-Riemannian geometry*, <https://webusers.imj-prg.fr/~davide.barilari/2016-11-21-ABB.pdf>, 2016.
- [ABGR09] Agrachev, A., Boscain, U., Gauthier, J.-P., and Rossi, F., *The intrinsic hypoelliptic Laplacian and its heat kernel on unimodular Lie groups*, J Funct Anal **256** (2009), no. 8, 2621–2655.
- [AF09] Astola, L. and Florack, L., *Finsler Geometry on Higher Order Tensor Fields and Applications to High Angular Resolution Diffusion Imaging*, Scale Space and Variational Methods in Computer Vision, Springer, Berlin, Heidelberg, June 2009, DOI: 10.1007/978-3-642-02256-2_19, pp. 224–234 (en).
- [AGLM93] Alvarez, L., Guichard, F., Lions, P.-L., and Morel, J.-M., *Axioms and fundamental equations of image processing*, Arch. Rational Mech. Anal. **123** (1993), no. 3, 199–257 (en).
- [AK06] Aubert, G. and Kornprobst, P., *Mathematical problems in image processing: partial differential equations and the calculus of variations*, vol. 147, Springer Science & Business Media, 2006.
- [ALJ+11] Aganj, I., Lenglet, C., Jahanshad, N., Yacoub, E., Harel, N., Thompson, P. M., and Sapiro, G., *A Hough transform global probabilistic approach*

- to multiple-subject diffusion MRI tractography, *Medical image analysis* **15** (2011), no. 4, 414–425.
- [Arz96] Arzela, C., *Sulle funzioni di linee*, no. 5, 225–244.
- [AS04] Agrachev, A. A. and Sachkov, Y. L., *Control Theory from the Geometric Viewpoint*, Encyclopaedia of Mathematical Sciences, vol. 87, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- [Asc83] Ascoli, G., *Le curvi limiti de una varietà data di curve*, no. 3, 521–586.
- [Aug01] August, J., *The Curve Indicator Random Field*, Ph.D. thesis, Yale University, 2001.
- [AZ03] August, J. and Zucker, S. W., *Sketches with curvature: the curve indicator random field and Markov processes*, *IEEE Trans Pattern Anal Mach Intell* **25** (2003), no. 4, 387–400.
- [BBN12] Barilari, D., Boscain, U., and Neel, R. W., *Small-time heat kernel asymptotics at the sub-Riemannian cut locus*, *J. Differential Geom.* **92** (2012), no. 3, 373–416 (EN). MR MR3005058
- [BBPW09] Burgeth, B., Breuß, M., Pizarro, L., and Weickert, J., *PDE-Driven Adaptive Morphology for Matrix Fields.*, SSVM, Lecture Notes in Computer Science, vol. 5567, Springer, May 2009, pp. 247–258.
- [BCD97] Bardi, M. and Capuzzo-Dolcetta, I., *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser Boston, Boston, MA, 1997 (en).
- [BCH⁺14] Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., Thompson, M. R., and Venkatesan, R., *Magnetic Resonance Imaging: Physical Principles and Sequence Design*, 2nd ed., Wiley-Blackwell, 2014.
- [BCP17] Bekkers, E. J., Chen, D., and Portegies, J. M., *Nilpotent Approximations of Sub-Riemannian Distances for Fast Perceptual Grouping of Blood Vessels in 2d and 3d*, arXiv:1707.02811 [math] (2017), arXiv: 1707.02811.
- [BCR10] Boscain, U., Charlot, G., and Rossi, F., *Existence of planar curves minimizing length and curvature*, *Proc. Steklov Inst. Math.* **270** (2010), no. 1, 43–56 (en).
- [BCS00] Bao, D., Chern, S.-S., and Shen, Z., *An Introduction to Riemann-Finsler Geometry*, Graduate Texts in Mathematics, vol. 200, Springer New York, New York, NY, 2000.

- [BDGR12] Boscaïn, U., Duplaix, J., Gauthier, J., and Rossi, F., *Anthropomorphic Image Reconstruction via Hypoelliptic Diffusion*, SIAM J. Control Optim. **50** (2012), no. 3, 1309–1336.
- [BDMS15] Bekkers, E., Duits, R., Mashtakov, A., and Sanguinetti, G., *A PDE Approach to Data-Driven Sub-Riemannian Geodesics in $SE(2)$* , SIAM J. Imaging Sci. **8** (2015), no. 4, 2740–2770.
- [BDMS17] Bekkers, E. J., Duits, R., Mashtakov, A., and Sachkov, Y., *Vessel Tracking via Sub-Riemannian Geodesics on the Projective Line Bundle*, Geometric Science of Information, Springer, Cham, November 2017, DOI: 10.1007/978-3-319-68445-1_89, pp. 773–781 (en).
- [BDRS14] Boscaïn, U., Duits, R., Rossi, F., and Sachkov, Y., *Curve cusplless reconstruction via sub-Riemannian geometry*, ESAIM: Control, Optimisation and Calculus of Variations **20** (2014), no. 3, 748–770.
- [Bek17] Bekkers, E. J., *Retinal image analysis using sub-Riemannian geometry in $SE(2)$* , Ph.D. thesis, Technische Universiteit Eindhoven, January 2017.
- [Bel96] Bellaïche, A., *The tangent space in sub-Riemannian geometry*, Sub-Riemannian Geometry, Progress in Mathematics, no. 144, Birkhäuser Basel, 1996, DOI: 10.1007/978-3-0348-9210-0_1, pp. 1–78 (en).
- [BLU07] Bonfiglioli, A., Lanconelli, E., and Uguzzoni, F., *Stratified Lie Groups and Potential Theory for their Sub-Laplacians*, Springer Monographs in Mathematics, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007 (en), DOI: 10.1007/978-3-540-71897-0.
- [BMLB94] Basser, P. J., Mattiello, J., and Le Bihan, D., *MR diffusion tensor spectroscopy and imaging*, Biophysical journal **66** (1994), no. 1, 259–267.
- [BPP⁺00] Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A., *In vivo fiber tractography using DT-MRI data*, Magn Reson Med **44** (2000), no. 4, 625–632 (eng).
- [CBG⁺12] Côté, M.-A., Boré, A., Girard, G., Houde, J.-C., and Descoteaux, M., *Tractometer: online evaluation system for tractography*, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012, Springer, 2012, pp. 699–706.
- [CDD⁺14] Caruyer, E., Daducci, A., Descoteaux, M., Houde, J.-C., Thiran, J.-P., and Verma, R., *Phantoms: a flexible software library to simulate diffusion MR phantoms*, ISMRM (Milan, Italy), May 2014.

- [CDDH11] Creusen, E. J., Duits, R., and Dela Haije, T. C. J., *Numerical Schemes for Linear and Non-linear Enhancement of DW-MRI*, Scale Space and Variational Methods in Computer Vision (Bruckstein, A. M., ter Haar Romeny, B. M., Bronstein, A. M., and Bronstein, M. M., eds.), LNCS, no. 6667, Springer Berlin Heidelberg, May 2011, pp. 14–25 (en).
- [CFSS16] Citti, G., Franceschiello, B., Sanguinetti, G., and Sarti, A., *Sub-Riemannian Mean Curvature Flow for Image Processing*, SIAM J. Imaging Sci. **9** (2016), no. 1, 212–237.
- [CGB+13] Côté, M.-A., Girard, G., Boré, A., Garyfallidis, E., Houde, J.-C., and Descoteaux, M., *Tractometer: Towards validation of tractography pipelines.*, Medical Image Analysis **17** (2013), no. 7, 844–857.
- [Che16] Chen, D., *New Minimal Path Model for Tubular Extraction and Image Segmentation*, phdthesis, PSL Research University, September 2016.
- [Chi11] Chirikjian, G. S., *Stochastic Models, Information Theory, and Lie Groups: Analytic Methods and Modern Applications*, vol. 2, Springer Science & Business Media, 2011.
- [CIL92] Crandall, M. G., Ishii, H., and Lions, P.-L., *User’s guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. **27** (1992), no. 1, 1–67.
- [CK97] Cohen, L. D. and Kimmel, R., *Global Minimum for Active Contour Models: A Minimal Path Approach*, International Journal of Computer Vision **24** (1997), no. 1, 57–78 (en).
- [CK00] Chirikjian, G. S. and Kyatkin, A. B., *Engineering Applications of Non-commutative Harmonic Analysis: With Emphasis on Rotation and Motion Groups*, CRC Press, September 2000 (en).
- [CKS97] Caselles, V., Kimmel, R., and Sapiro, G., *Geodesic Active Contours*, International Journal of Computer Vision **22** (1997), no. 1, 61–79 (en).
- [CL83] Crandall, M. G. and Lions, P.-L., *Viscosity solutions of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc. **277** (1983), no. 1, 1–42.
- [CMC16a] Chen, D., Mirebeau, J.-M., and Cohen, L. D., *Global Minimum for a Finsler Elastica Minimal Path Approach*, Int J Comput Vis (2016), 1–26 (en).

- [CMC16b] ———, *Vessel tree extraction using radius-lifted keypoints searching scheme and anisotropic fast marching method*, Journal of Algorithms & Computational Technology **10** (2016), no. 4, 224–234 (en).
- [CP11] Chambolle, A. and Pock, T., *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision **40** (2011), no. 1, 120–145.
- [CS05] Chan, T. and Shen, J., *Image Processing and Analysis*, Other Titles in Applied Mathematics, Society for Industrial and Applied Mathematics, January 2005.
- [CS06] Citti, G. and Sarti, A., *A Cortical Based Model of Perceptual Completion in the Roto-Translation Space*, J Math Imaging Vis **24** (2006), no. 3, 307–326 (en).
- [CTC⁺09] Close, T. G., Tournier, J.-D., Calamante, F., Johnston, L. A., Mareels, I., and Connelly, A., *A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms*, NeuroImage **47** (2009), no. 4, 1288–1300.
- [CTJC10] Calamante, F., Tournier, J.-D., Jackson, G. D., and Connelly, A., *Track-density imaging (TDI): super-resolution white matter imaging using whole-brain track-density mapping*, Neuroimage **53** (2010), no. 4, 1233–1243.
- [CWF⁺14] Chamberland, M., Whittingstall, K., Fortin, D., Mathieu, D., and Descoteaux, M., *Real-time multi-peak tractography for instantaneous connectivity display*, Front. Neuroinform. **8** (2014) (English).
- [CXB⁺15] Cauley, S. F., Xi, Y., Bilgic, B., Xia, J., Adalsteinsson, E., Balakrishnan, V., Wald, L. L., and Setsompop, K., *Fast reconstruction for multichannel compressed sensing using a hierarchically semiseparable solver*, Magnetic Resonance in Medicine **73** (2015), 1034–1040.
- [DB07] Duits, R. and Burgeth, B., *Scale Spaces on Lie Groups.*, SSVM, Lecture Notes in Computer Science, vol. 4485, Springer, August 2007, pp. 300–312.
- [DBRS13] Duits, R., Boscain, U., Rossi, F., and Sachkov, Y., *Association Fields via Cuspless Sub-Riemannian Geodesics in $SE(2)$* , J Math Imaging Vis **49** (2013), no. 2, 384–417 (en).
- [DCDT13] Daducci, A., Caruyer, E., Descoteaux, M., and Thiran, J.-P., *HARDI Reconstruction Challenge*, 2013, Published: IEEE International Symposium on Biomedical Imaging.

- [DCRD⁺14] Daducci, A., Canales-Rodríguez, E. J., Descoteaux, M., Garyfallidis, E., Gur, Y., Lin, Y.-C., Mani, M., Merlet, S., Paquette, M., Ramirez-Manzanares, A., Reisert, M., Rodrigues, P. R., Sepelband, F., Jacob, M., Caruyer, E., Choupan, J., Deriche, R., Menegaz, G., Prčkovska, V., Rivera, M., Wiaux, Y., and Thiran, J.-P., *Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI*, IEEE Transactions on Medical Imaging **33** (2014), no. 2, 384–399.
- [DDHCG12] Duits, R., Dela Haije, T., Creusen, E., and Ghosh, A., *Morphological and Linear Scale Spaces for Fiber Enhancement in DW-MRI*, J Math Imaging Vis **46** (2012), no. 3, 326–368 (en).
- [DDKA09] Descoteaux, M., Deriche, R., Knosche, T. R., and Anwander, A., *Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions*, IEEE Transactions on Medical Imaging **28** (2009), no. 2, 269–286.
- [DDL^B+11] Descoteaux, M., Deriche, R., Le Bihan, D., Mangin, J.-F., and Poupon, C., *Multiple q -shell diffusion propagator imaging*, Med Image Anal **15** (2011), no. 4, 603–621 (eng).
- [DEVH⁺14] Dhollander, T., Emsell, L., Van Hecke, W., Maes, F., Sunaert, S., and Suetens, P., *Track orientation density imaging (TODI) and track orientation distribution (TOD) based tractography*, NeuroImage **94** (2014), 312–336.
- [DF09] Duits, R. and Franken, E., *Line Enhancement and Completion via Linear Left Invariant Scale Spaces on $SE(2)$* , SSVM (Berlin, Heidelberg), Springer-Verlag, 2009, pp. 795–807.
- [DF10a] ———, *Left-Invariant Diffusions on the Space of Positions and Orientations and their Application to Crossing-Preserving Smoothing of HARDI images*, Int J Comput Vis **92** (2010), no. 3, 231–264 (en).
- [DF10b] ———, *Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores Part I: Linear left-invariant diffusion equations on $SE(2)$* , Quart. Appl. Math. **68** (2010), no. 2, 255–292.
- [DF10c] ———, *Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores Part II: Nonlinear left-invariant diffusions on invertible orientation scores*, Quart. Appl. Math. **68** (2010), no. 2, 293–331.

- [DF11] ———, *Left-Invariant Diffusions on the Space of Positions and Orientations and their Application to Crossing-Preserving Smoothing of HARDI images.*, International Journal of Computer Vision **92** (2011), no. 3, 231–264.
- [DFGR04] Duits, R., Florack, L., Graaf, J. d., and Romeny, B. t. H., *On the Axioms of Scale Space Theory*, Journal of Mathematical Imaging and Vision **20** (2004), no. 3, 267–298 (en).
- [DFGtHR06] Duits, R., Felsberg, M., Granlund, G., and ter Haar Romeny, B., *Image Analysis and Reconstruction using a Wavelet Transform Constructed from a Reducible Representation of the Euclidean Motion Group*, Int J Comput Vision **72** (2006), no. 1, 79–102 (en).
- [DGDHM16] Duits, R., Ghosh, A., Dela Haije, T. C. J., and Mashtakov, A., *On Sub-Riemannian Geodesics in $SE(3)$ Whose Spatial Projections do not Have Cusps*, J Dyn Control Syst **22** (2016), no. 4, 771–805 (en).
- [DGDHS14] Duits, R., Ghosh, A., Dela Haije, T., and Sachkov, Y., *Cuspless Sub-Riemannian Geodesics within the Euclidean Motion Group $SE(d)$* , Neuro-mathematics of Vision, Lecture Notes in Morphogenesis, 2014, DOI: 10.1007/978-3-642-34444-2_5, pp. 173–215 (en).
- [DH94] Driscoll, J. R. and Healy, D. M., *Computing Fourier Transforms and Convolutions on the 2-Sphere*, Advances in Applied Mathematics **15** (1994), no. 2, 202 – 250.
- [DHDT14] Dela Haije, T., Duits, R., and Tax, C., *Sharpening fibers in diffusion weighted MRI via erosion*, Visualization and processing of tensors and higher order descriptors for multi-valued data (Westin, C., Vilanova, A., and Burgeth, B., eds.), Mathematics and visualization, 2014.
- [DJHS15] Duits, R., Janssen, M., Hannink, J., and Sanguinetti, G., *Locally adaptive frames in the roto-translation group and their applications in medical imaging*, arXiv:1502.08002v4 (2015).
- [DJHS16] Duits, R., Janssen, M. H. J., Hannink, J., and Sanguinetti, G. R., *Locally Adaptive Frames in the Roto-Translation Group and Their Applications in Medical Imaging*, J Math Imaging Vis (2016), 1–36 (en).
- [DMMP16] Duits, R., Meesters, S. P. L., Mirebeau, J.-M., and Portegies, J. M., *Optimal Paths for Variants of the 2d and 3d Reeds-Shepp Car with*

- Applications in Image Analysis*, arXiv:1612.06137 [math] (2016), arXiv:1612.06137.
- [DP12] Descoteaux, M. and Poupon, C., *Diffusion-Weighted MRI*, Comprehensive Biomedical Physics, Elsevier, 2012.
- [Dra05] Dragoni, F., *Carnot-Carathéodory Metrics and Viscosity Solutions*, Ph.D. thesis, Scuola Normale Superiore di Pisa, 2005.
- [Dra07] ———, *Metric Hopf-Lax formula with semicontinuous data*, Discrete & Continuous Dynamical Systems **17** (2007), no. 4, 713–729 (en).
- [Dub57] Dubins, L. E., *On Curves of Minimal Length with a Constraint on Average Curvature, and with Prescribed Initial and Terminal Positions and Tangents*, American Journal of Mathematics **79** (1957), no. 3, 497–516.
- [Dui05] Duits, R., *Perceptual organization in image analysis*, Ph.D. thesis, Technische Universiteit Eindhoven, 2005.
- [DvA08] Duits, R. and van Almsick, M., *The explicit solutions of linear left-invariant second order stochastic evolution equations on the 2d Euclidean motion group*, Quart. Appl. Math. **66** (2008), no. 1, 27–67.
- [Eva10] Evans, L. C., *Partial Differential Equations*, 2 edition ed., American Mathematical Society, Providence, R.I, March 2010 (English).
- [FD09a] Franken, E. and Duits, R., *Crossing-Preserving Coherence-Enhancing Diffusion on Invertible Orientation Scores*, Int J Comput Vis **85** (2009), no. 3, 253–278 (en).
- [FD09b] Franken, E. and Duits, R., *Crossing-preserving coherence-enhancing diffusion on invertible orientation scores*, International Journal of Computer Vision **85** (2009), no. 3, 253–278.
- [FDB04] Florack, L., Duits, R., and Bierkens, J., *Tikhonov regularization versus scale space: a new result*, International Conference on Image Processing, vol. 1, October 2004, pp. 271–274 Vol. 1.
- [FGB⁺14] Feng, L., Grimm, R., Block, K. T., Chandarana, H., Kim, S., Xu, J., Axel, L., Sodickson, D. K., and Otazo, R., *Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI*, Magnetic Resonance in Medicine **72** (2014), 707–717.

- [Fig07] Figueiredo, B. D. B., *Generalized spheroidal wave equation and limiting cases*, J Math Phys **48** (2007), no. 1, 013503.
- [Fis12] Fischl, B., *FreeSurfer*, NeuroImage **62** (2012), no. 2, 774–781.
- [FJ07] Fletcher, P. T. and Joshi, S., *Riemannian geometry for the statistical analysis of diffusion tensor data*, Signal Processing **87** (2007), no. 2, 250 – 262, Tensor Signal Processing.
- [Fla57] Flammer, C., *Spheroidal Wave Functions*, Stanford University Press, 1957 (en).
- [Flo97] Florack, L., *Image Structure*, Computational Imaging and Vision, Springer, Dordrecht, 1997 (en), DOI: 10.1007/978-94-015-8845-4_1.
- [FM14] Fehrenbach, J. and Mirebeau, J.-M., *Sparse Non-negative Stencils for Anisotropic Diffusion*, J Math Imaging Vis **49** (2014), no. 1, 123–147 (en).
- [FMS⁺10] Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., Miller, K. L., Ugurbil, K., and Yacoub, E., *Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging*, PLoS ONE **5** (2010), no. 12.
- [Fol73] Folland, G. B., *A fundamental solution for a subelliptic operator*, Bull. Amer. Math. Soc. **79** (1973), no. 2, 373–376 (EN). MR MR0315267
- [Fol94] Folland, G. B., *A Course in Abstract Harmonic Analysis*, CRC Press, December 1994 (en).
- [Fra08] Franken, E., *Enhancement of crossing elongated structures in images*, Ph.D. thesis, Eindhoven University of Technology, Department of Biomedical Engineering, The Netherlands, 2008.
- [FS63] Falconer, M. A. and Serafetinides, E. A., *A follow-up study of surgery in temporal lobe epilepsy*, Journal of neurology, neurosurgery, and psychiatry **26** (1963), no. 2, 154.
- [FS04] Felsberg, M. and Sommer, G., *The Monogenic Scale-Space: A Unifying Approach to Phase-Based Image Processing in Scale-Space*, Journal of Mathematical Imaging and Vision **21** (2004), no. 1-2, 5–26 (en).
- [Füh05] Führ, H., *Abstract Harmonic Analysis of Continuous Wavelet Transforms*, Springer Science & Business Media, 2005 (en).

- [GBA⁺14] Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., and Nimmo-Smith, I., *Dipy, a library for the analysis of diffusion MRI data*, *Frontiers in Neuroinformatics* **8** (2014), no. 8.
- [GDS⁺15] Golkov, V., Dosovitskiy, A., Sämann, P., Sperl, J. I., Sprenger, T., Czisch, M., Menzel, M. I., Gómez, P. A., Haase, A., Brox, T., and Cremers, D., *q-Space deep learning for twelve-fold shorter and model-free diffusion MRI scans*, *MICCAI*, 2015.
- [GPG⁺16] Golkov, V., Portegies, J. M., Golkov, A., Duits, R., and Cremers, D., *Holistic Image Reconstruction for Diffusion MRI*, *Computational Diffusion MRI*, Springer, Cham, 2016, DOI: 10.1007/978-3-319-28588-7_3, pp. 27–39 (en).
- [Gro96] Gromov, M., *Carnot-Carathéodory spaces seen from within*, *Sub-Riemannian Geometry*, *Progress in Mathematics*, no. 144, Birkhäuser Basel, 1996, DOI: 10.1007/978-3-0348-9210-0_2, pp. 79–323 (en).
- [GSM⁺15] Golkov, V., Sperl, J. I., Menzel, M. I., Sprenger, T., Tan, E. T., Marinelli, L., Hardy, C. J., Haase, A., and Cremers, D., *Joint super-resolution using only one anisotropic low-resolution image per q-space coordinate*, *Computational Diffusion MRI*, *MICCAI Workshop 2014* (O’Donnell, L., Nedjati-Gilani, G., Rathi, Y., Reisert, M., and Schneider, T., eds.), Springer International Publishing, 2015, pp. 181–191.
- [GSW⁺13] Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., and Jenkinson, M., *The minimal preprocessing pipelines for the Human Connectome Project*, *NeuroImage* **80** (2013), 105–124.
- [GWDD14] Girard, G., Whittingstall, K., Deriche, R., and Descoteaux, M., *Towards quantitative connectivity analysis: reducing tractography biases*, *NeuroImage* **98** (2014), no. 0, 266 – 278.
- [HJM⁺06] Hagmann, P., Jonasson, L., Maeder, P., Thiran, J.-P., Wedeen, V. J., and Meuli, R., *Understanding Diffusion MR Imaging Techniques: From Scalar Diffusion-weighted Imaging to Diffusion Tensor Imaging and Beyond*, *RadioGraphics* **26** (2006), no. suppl_1, S205–S223.
- [Hör67] Hörmander, L., *Hypoelliptic second order differential equations*, *Acta Mathematica* **119** (1967), no. 1, 147–171 (English).

- [Iij62] Iijima, T., *Observation theory of two-dimensional visual patterns*, Technical Group on Automata and Automatic Control, IECE, Japan (1962).
- [JBB⁺12] Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M., *FSL*, *NeuroImage* **62** (2012), no. 2, 782–790.
- [JBBS02] Jenkinson, M., Bannister, P., Brady, M., and Smith, S., *Improved optimization for the robust and accurate linear registration and motion correction of brain images*, *Neuroimage* **17** (2002), no. 2, 825–841 (eng).
- [JBP98] Jezzard, P., Barnett, A. S., and Pierpaoli, C., *Characterization of and correction for eddy current artifacts in echo planar diffusion imaging*, *Magn. Reson. Med.* **39** (1998), no. 5, 801–812 (en).
- [JBT⁺08] Jbabdi, S., Bellec, P., Toro, R., Daunizeau, J., Pélégrini-Issac, M., and Bernali, H., *Accurate Anisotropic Fast Marching for Diffusion-Based Geodesic Tractography*, *Int J Biomed Imaging* **2008** (2008).
- [JDB15] Janssen, M., Duits, R., and Breeuwer, M., *Invertible Orientation Scores of 3d Images*, *Scale Space and Variational Methods in Computer Vision*, Springer, Cham, May 2015, DOI: 10.1007/978-3-319-18461-6_45, pp. 563–575 (en).
- [JHR⁺05] Jensen, J. H., Helpert, J. A., Ramani, A., Lu, H., and Kaczynski, K., *Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging*, *Magn. Reson. Med.* **53** (2005), no. 6, 1432–1440 (en).
- [JJB11] Jbabdi, S. and Johansen-Berg, H., *Tractography: Where Do We Go from Here?*, *Brain Connectivity* **1** (2011), no. 3, 169–183.
- [JLT⁺13] Jeurissen, B., Leemans, A., Tournier, J.-D., Jones, D. K., and Sijbers, J., *Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging*, *Human Brain Mapping* **34** (2013), no. 11, 2747–2766.
- [JS86] Jerison, D. S. and Sánchez-Calle, A., *Estimates for the Heat Kernel for a Sum of Squares of Vector Fields*, *Indiana University Mathematics Journal* **35** (1986), no. 4, 835–854.
- [JTD⁺14a] Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., and Sijbers, J., *Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data*, *NeuroImage* **103** (2014), 411–426.

- [JTD⁺14b] ———, *Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data*, *Neuroimage* **103** (2014), 411–426 (eng).
- [Kap80] Kaplan, A., *Fundamental Solutions for a Class of Hypoelliptic PDE Generated by Composition of Quadratic Forms*, *Transactions of the American Mathematical Society* **258** (1980), no. 1, 147–153.
- [Kat76] Kato, T., *Operators in Hilbert spaces*, *Perturbation Theory for Linear Operators*, *Classics in Mathematics*, Springer Berlin Heidelberg, 1976, pp. 251–308 (en).
- [KHK⁺12] Khare, K., Hardy, C. J., King, K. F., Turski, P. A., and Marinelli, L., *Accelerated MR imaging using compressive sensing with no free parameters*, *Magnetic Resonance in Medicine* **68** (2012), no. 5, 1450–7.
- [Kna02] Knapp, A., *Lie Groups Beyond an Introduction*, 2002.
- [Koe84] Koenderink, J. J., *The structure of images*, *Biol. Cybern.* **50** (1984), no. 5, 363–370 (en).
- [Kor82] Korányi, A., *Kelvin transforms and harmonic polynomials on the Heisenberg group*, *Journal of Functional Analysis* **49** (1982), no. 2, 177–185.
- [KRV99] Kalitzin, S. N., Romeny, B. M. T. H., and Viergever, M. A., *Invertible Apertured Orientation Filters in Image Analysis*, *International Journal of Computer Vision* **31** (1999), no. 2-3, 145–158 (en).
- [LBBL⁺86] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., and Laval-Jeantet, M., *MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders.*, *Radiology* **161** (1986), no. 2, 401–407.
- [LDP07] Lustig, M., Donoho, D., and Pauly, J. M., *Sparse MRI: The application of compressed sensing for rapid MR imaging.*, *Magnetic Resonance in Medicine* **58** (2007), no. 6, 1182–95.
- [Lea86] Leaver, E. W., *Solutions to a generalized spheroidal wave equation: Teukolsky's equations in general relativity, and the two-center problem in molecular quantum mechanics*, *J Math Phys* **27** (1986), no. 5, 1238–1265.
- [Lin94] Lindeberg, T., *Scale-Space Theory in Computer Vision*, *The Springer International Series in Engineering and Computer Science*, Springer, Boston, MA, 1994 (en), DOI: 10.1007/978-1-4757-6465-9_1.

- [LPP⁺09] Lenglet, C., Prados, E., Pons, J., Deriche, R., and Faugeras, O., *Brain Connectivity Mapping Using Riemannian Geometry, Control Theory, and PDEs*, SIAM J. Imaging Sci. **2** (2009), no. 2, 285–322.
- [LWC⁺03] Lin, C. P., Wedeen, V. J., Chen, J. H., Yao, C., and Tseng, W. Y. I., *Validation of diffusion spectrum magnetic resonance imaging with manganese-enhanced rat optic tracts and ex vivo phantoms*, NeuroImage **19** (2003), 482–495.
- [Mac49] Mackey, G. W., *Imprimitivity for Representations of Locally Compact Groups I*, Proc Natl Acad Sci U S A **35** (1949), no. 9, 537–545.
- [Mak12] Makin, A. S., *On Summability of Spectral Expansions Corresponding to the Sturm-Liouville Operator*, Int J Math Math Sci **2012**, **2012** (2012), e843562 (en).
- [MAS13] Mashtakov, A. P., Ardentov, A. A., and Sachkov, Y. L., *Parallel Algorithm and Software for Image inpainting via Sub-Riemannian Minimizers on the Group of Rototranslations*, Numerical Mathematics: Theory, Methods and Applications **6** (2013), no. 01, 95–115.
- [MCCVZ99] Mori, S., Crain, B. J., Chacko, V. P., and Van Zijl, P. C. M., *Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging*, Ann Neurol. **45** (1999), no. 2, 265–269 (en).
- [MDS⁺17] Mashtakov, A., Duits, R., Sachkov, Y. L., Bekkers, E. J., and Beschastnyi, I., *Tracking of Lines in Spherical Images via Sub-Riemannian Geodesics in $SO(3)$* , J Math Imaging Vis (2017), 1–26 (en).
- [Mee13] Meesters, S., *Diffusion weighted tractography to reconstruct the optic radiation in support of temporal lobe epilepsy surgery*, Ph.D. thesis, Eindhoven University of Technology, 2013.
- [MHNH⁺17] Maier-Hein, K. H., Neher, P. F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Stieltjes, B., and Descoteaux, M., *The challenge of mapping the human connectome based on diffusion tractography*, Nature Communications **8** (2017), no. 1, 1349 (En).
- [Mir13] Mirebeau, J.-M., *Efficient fast marching with Finsler metrics*, Numer. Math. **126** (2013), no. 3, 515–557 (en).
- [Mir14] ———, *Anisotropic Fast-Marching on Cartesian Grids Using Lattice Basis Reduction*, SIAM J. Numer. Anal. **52** (2014), no. 4, 1573–1599.

- [Mir17a] Mirebeau, J.-M., *Anisotropic fast-marching on cartesian grids using Voronoi's first reduction of quadratic forms* (en).
- [Mir17b] ———, *Fast Marching methods for Curvature Penalized Shortest Paths*, working paper or preprint, June 2017.
- [MJG⁺15] Mani, M., Jacob, M., Guidon, A., Magnotta, V., and Zhong, J., *Acceleration of high angular and spatial resolution diffusion imaging using compressed sensing with multichannel spiral data*, *Magnetic Resonance in Medicine* **73** (2015), 126–138.
- [MM56] Margenau, H. and Murphy, G. M., *The mathematics of physics and chemistry*, Van Nostrand, 1956 (en).
- [MMN⁺07] Melonakos, J., Mohan, V., Niethammer, M., Smith, K., Kubicki, M., and Tannenbaum, A., *Finsler Tractography for White Matter Connectivity Analysis of the Cingulum Bundle*, *Med Image Comput Comput Assist Interv* **10** (2007), no. 0 1, 36–43.
- [Mon02] Montgomery, R., *A Tour of Subriemannian Geometries, Their Geodesics and Applications*, *Mathematical Surveys and Monographs*, vol. 91, 2002.
- [MPAT08] Melonakos, J., Pichon, E., Angenent, S., and Tannenbaum, A., *Finsler active contours*, *IEEE Trans Pattern Anal Mach Intell* **30** (2008), no. 3, 412–423 (eng).
- [MS54] Meixner, D. J. and Schäfke, D. F. W., *Grundlagen*, *Mathematische Funktionen und Sphäroidfunktionen, Die Grundlehren der Mathematischen Wissenschaften*, no. 71, Springer Berlin Heidelberg, 1954, DOI: 10.1007/978-3-662-00941-3_2, pp. 14–98 (de).
- [MS10] Moiseev, I. and Sachkov, Y. L., *Maxwell strata in sub-Riemannian problem on the group of motions of a plane*, *ESAIM: Control, Optimisation and Calculus of Variations* **16** (2010), no. 02, 380–399.
- [MSG⁺16a] Meesters, S. P. L., Sanguinetti, G. R., Garyfallidis, E., Portegies, J. M., and Duits, R., *Fast implementations of contextual PDE's for HARDI data processing in DIPY*, 2016.
- [MSG⁺16b] Meesters, S.P.L., Sanguinetti, G. R., Garyfallidis, E., Portegies, J. M., Ossenblok, P., and Duits, R., *Cleaning output of tractography via fiber to bundle coherence, a new open source implementation*, June 2016.

- [MSS09] Momayyez-Siahkal, P. and Siddiqi, K., *3d Stochastic Completion Fields for Fiber Tractography*, Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, June 2009, pp. 178–185.
- [MSS13] ———, *3d Stochastic Completion Fields for Mapping Connectivity in Diffusion MRI*, IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013), no. 4, 983–995.
- [Mum94] Mumford, D., *Elastica and Computer Vision*, Algebraic Geometry and its Applications (Bajaj, C. L., ed.), Springer New York, 1994, pp. 491–506 (en).
- [NSL⁺07] Nilsson, D., Starck, G., Ljungberg, M., Ribbelin, S., Jönsson, L., Malmgren, K., and Rydenhag, B., *Intersubject variability in the anterior extent of the optic radiation assessed by tractography*, Epilepsy research **77** (2007), no. 1, 11–16.
- [NSW85] Nagel, A., Stein, E. M., and Wainger, S., *Balls and metrics defined by vector fields I: Basic properties*, Acta Math. **155** (1985), 103–147 (EN).
- [OLBC10] Olver, F. W., Lozier, D. W., Boisvert, R. F., and Clark, C. W., *NIST Handbook of Mathematical Functions*, 1st ed., Cambridge University Press, 2010.
- [OVM05] Ozarslan, E., Vemuri, B. C., and Mareci, T. H., *Generalized scalar measures for diffusion MRI using trace, variance, and entropy*, Magn Reson Med **53** (2005), no. 4, 866–876 (eng).
- [PAV⁺15] Prčkovska, V., Andorrà, M., Villoslada, P., Martinez-Heras, E., Duits, R., Fortin, D., Rodrigues, P., and Descoteaux, M., *Contextual Diffusion Image Post-processing Aids Clinical Applications*, Visualization and Processing of Higher Order Descriptors for Multi-Valued Data (Hotz, I. and Schultz, T., eds.), Mathematics and Visualization, Springer International Publishing, 2015, DOI: 10.1007/978-3-319-15090-1_18, pp. 353–377 (en).
- [PB96] Pierpaoli, C. and Basser, P. J., *Toward a quantitative assessment of diffusion anisotropy*, Magn. Reson. Med. **36** (1996), no. 6, 893–906 (en).
- [PB14] Parikh, N. and Boyd, S., *Proximal Algorithms*, Foundations and Trends in Optimization **1** (2014), no. 3, 123–231.
- [PBG15] Prandi, D., Boscaïn, U., and Gauthier, J.-P., *Image Processing in the Semidiscrete Group of Rototranslations*, Geometric Science of Information (Nielsen, F. and Barbaresco, F., eds.), Lecture Notes in Computer

- Science, no. 9389, Springer International Publishing, October 2015, DOI: 10.1007/978-3-319-25040-3_67, pp. 627–634 (en).
- [PCBC09] Pock, T., Cremers, D., Bischof, H., and Chambolle, A., *An algorithm for minimizing the Mumford–Shah functional*, 2009 IEEE 12th International Conference on Computer Vision (ICCV), IEEE, September 2009, pp. 1133–1140.
- [PD17] Portegies, J. M. and Duits, R., *New exact and numerical solutions of the (convection–)diffusion kernels on $SE(3)$* , Differential Geometry and its Applications **53** (2017), 182–219.
- [PDK09] Péchaud, M., Descoteaux, M., and Keriven, R., *Brain connectivity using geodesics in HARDI*, Med Image Comput Comput Assist Interv **12** (2009), no. Pt 2, 482–489 (eng).
- [Pet03] Petitot, J., *The neurogeometry of pinwheels as a sub-Riemannian contact structure*, Journal of Physiology-Paris **97** (2003), no. 2–3, 265–309.
- [PFS⁺15] Portegies, J. M., Fick, R. H. J., Sanguinetti, G. R., Meesters, S. P. L., Girard, G., and Duits, R., *Improving Fiber Alignment in HARDI by Combining Contextual PDE Flow with Constrained Spherical Deconvolution*, PLoS ONE **10** (2015), no. 10, e0138122.
- [PGFM95] Pauwels, E. J., Gool, L. J. v., Fiddelaers, P., and Moons, T., *An extended class of scale-invariant and recursive scale space filters*, IEEE Transactions on Pattern Analysis and Machine Intelligence **17** (1995), no. 7, 691–701.
- [PJB⁺13] Poot, D. H. J., Jeurissen, B., Bastiaensen, Y., Veraart, J., Van Hecke, W., Parizel, P. M., and Sijbers, J., *Super-resolution for multislice diffusion tensor imaging*, Magnetic Resonance in Medicine **69** (2013), no. 1, 103–113.
- [PM90] Perona, P. and Malik, J., *Scale-Space and Edge Detection Using Anisotropic Diffusion*, IEEE Trans. Pattern Anal. Mach. Intell. **12** (1990), no. 7, 629–639.
- [PMG⁺15] Paquette, M., Merlet, S., Gilbert, G., Deriche, R., and Descoteaux, M., *Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging.*, Magnetic Resonance in Medicine **73** (2015), 401–416.
- [PPA⁺05] Powell, H., Parker, G., Alexander, D., Symms, M., Boulby, P., Wheeler-Kingshott, C., Barker, G., Koeppe, M., and Duncan, J., *MR tractography*

- predicts visual field defects following temporal lobe resection*, *Neurology* **65** (2005), no. 4, 596–599.
- [PPKC10] Peyré, G., Péchaud, M., Keriven, R., and Cohen, L. D., *Geodesic Methods in Computer Vision and Graphics*, *Found. Trends. Comput. Graph. Vis.* **5** (2010), no. 3–4, 197–397.
- [PRD⁺10] Prčkovska, V., Rodrigues, P., Duits, R., Haar Romenij, B. t., and Vilanova, A., *Extrapolating fiber crossings from DTI data: can we infer similar fiber crossings as in HARDI?*, MICCAI, Workshop on Computational Diffusion MRI, 2010.
- [PSMD15] Portegies, J., Sanguinetti, G., Meesters, S., and Duits, R., *New Approximation of a Scale Space Kernel on $SE(3)$ and Applications in Neuroimaging*, SSVN (Aujol, J.-F., Nikolova, M., and Papadakis, N., eds.), LNCS, no. 9087, Springer International Publishing, May 2015, pp. 40–52 (en).
- [RDtHRV10] Rodrigues, P., Duits, R., ter Haar Romeny, B. M., and Vilanova, A., *Accelerated Diffusion Operators for Enhancing DW-MRI*, Proceedings of the 2Nd Eurographics Conference on Visual Computing for Biology and Medicine (Aire-la-Ville, Switzerland, Switzerland), EG VCBM'10, Eurographics Association, 2010, pp. 49–56.
- [RJP⁺14] Roine, T., Jeurissen, B., Perrone, D., Aelterman, J., Leemans, A., Philips, W., and Sijbers, J., *Isotropic non-white matter partial volume effects in constrained spherical deconvolution*, *Frontiers in Neuroinformatics* **8** (2014), no. 28.
- [RJP⁺15] Roine, T., Jeurissen, B., Perrone, D., Aelterman, J., Philips, W., Leemans, A., and Sijbers, J., *Informed constrained spherical deconvolution (iCSD)*, *Medical Image Analysis* (2015), no. 0, In press.
- [RKK12] Reisert, M., Kellner, E., and Kiselev, V. G., *About the Geometry of Asymmetric Fiber Orientation Distributions*, *IEEE Trans Med Imaging* **31** (2012), no. 6, 1240–1249.
- [RML⁺14] Rathi, Y., Michailovich, O., Laun, F., Setsompop, K., Grant, P. E., and Westin, C.-F., *Multi-shell diffusion signal recovery from sparse measurements.*, *Medical Image Analysis* **18** (2014), no. 7, 1143–1156.
- [Ros06] Rossmann, W., *Lie Groups: An Introduction through Linear Groups*, Oxford University Press, Oxford, August 2006 (English).

- [RS76] Rothschild, L. P. and Stein, E. M., *Hypoelliptic differential operators and nilpotent groups*, Acta Math. **137** (1976), 247–320 (EN).
- [RS90] Reeds, J. A. and Shepp, L. A., *Optimal paths for a car that goes both forwards and backwards.*, Pacific J. Math. **145** (1990), no. 2, 367–393. MR MR1069892
- [RS12] Reisert, M. and Skibbe, H., *Left-Invariant Diffusion on the Motion Group in terms of the Irreducible Representations of $SO(3)$* , Computing Research Repository [abs/1202.5414](#) (2012).
- [RS13] ———, *Fiber Continuity Based Spherical Deconvolution in Spherical Harmonic Domain.*, MICCAI (3) (Mori, K., Sakuma, I., Sato, Y., Barillot, C., and Navab, N., eds.), Lecture Notes in Computer Science, vol. 8151, Springer, 2013, pp. 493–500.
- [RT92] Rouy, E. and Tourin, A., *A Viscosity Solutions Approach to Shape-From-Shading*, SIAM J. Numer. Anal. **29** (1992), no. 3, 867–884.
- [RTB⁺17] Rokem, A., Takemura, H., Bock, A. S., Scherf, K. S., Behrmann, M., Wandell, B. A., Fine, I., Bridge, H., and Pestilli, F., *The visual white matter: The application of diffusion MRI and fiber tractography to vision science* Rokem et al., Journal of Vision **17** (2017), no. 2, 4–4.
- [Rud06] Rudin, W., *Functional Analysis*, McGraw-Hill, 2006 (en).
- [Sac11] Sachkov, Y. L., *Cut locus and optimal synthesis in the sub-Riemannian problem on the group of motions of a plane*, ESAIM: Control, Optimisation and Calculus of Variations **17** (2011), no. 2, 293–321.
- [San11] Sanguinetti, G. R., *Invariant models of vision between phenomenology, image statistics and neurosciences*, Ph.D. thesis, Universidad de la República, Facultad de Ingeniería, 2011.
- [Sap06] Sapiro, G., *Geometric Partial Differential Equations and Image Analysis*, Cambridge University Press, New York, NY, USA, 2006.
- [SBD⁺15] Sanguinetti, G. R., Bekkers, E. J., Duits, R., Janssen, M. H. J., Mashnikov, A., and Mirebeau, J.-M., *Sub-Riemannian Fast Marching in $SE(2)$* , Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Lecture Notes in Computer Science, no. 9423, Springer International Publishing, 2015, DOI: 10.1007/978-3-319-25751-8_44, pp. 366–374 (en).

- [SBG⁺14] Stefani, G., Boscain, U., Gauthier, J.-P., Sarychev, A., and Sigalotti, M. (eds.), *Geometric Control Theory and Sub-Riemannian Geometry*, Springer INdAM Series, vol. 5, Springer International Publishing, Cham, 2014.
- [SCAG⁺12] Setsompop, K., Cohen-Adad, J., Gagoski, B. A., Raij, T., Yendiki, A., Keil, B., Wedeen, V. J., and Wald, L. L., *Improving diffusion MRI using simultaneous multi-slice echo planar imaging.*, *NeuroImage* **63** (2012), no. 1, 569–80.
- [SD15] Sharma, U. and Duits, R., *Left-invariant evolutions of wavelet transforms on the similitude group*, *Appl Comput Harmon Anal* **39** (2015), no. 1, 110–137.
- [SDBS⁺08] Sherbondy, A. J., Dougherty, R. F., Ben-Shachar, M., Napel, S., and Wandell, B. A., *ConTrack: finding the most likely pathways between brain regions using diffusion tractography*, *Journal of Vision* **8** (2008), no. 9, 15.1–16.
- [Sep11] Sepasian, N., *Multi-valued geodesic tractography for diffusion weighted imaging*, Ph.D. thesis, Technische Universiteit Eindhoven, 2011.
- [Set99] Sethian, J. A., *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, Cambridge University Press, June 1999 (en), Google-Books-ID: ErpOoynE4dIC.
- [SG13] Schultz, T. and Groeschel, S., *Auto-calibrating spherical deconvolution based on ODF sparsity*, MICCAI 2013, Springer, 2013, pp. 663–670.
- [SGW12] Scherrer, B., Gholipour, A., and Warfield, S. K., *Super-resolution reconstruction to increase the spatial resolution of diffusion weighted images from orthogonal anisotropic acquisitions.*, *Medical Image Analysis* **16** (2012), no. 7, 1465–76.
- [SJ95] Söderman, O. and Jönsson, B., *Restricted Diffusion in Cylindrical Geometry*, *Journal of Magnetic Resonance, Series A* **117** (1995), no. 1, 94 – 97.
- [SJX⁺13] Sotiropoulos, S. N., Jbabdi, S., Xu, J., Andersson, J. L., Moeller, S., Auerbach, E. J., Glasser, M. F., Hernandez, M., Sapiro, G., Jenkinson, M., Feinberg, D. a., Yacoub, E., Lenglet, C., Van Essen, D. C., Ugurbil, K., and Behrens, T. E. J., *Advances in diffusion MRI acquisition and processing in the Human Connectome Project.*, *NeuroImage* **80** (2013), 125–43.

- [ST65] Stejskal, E. and Tanner, J., *Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient*, Journal of Chemical Physics **42** (1965), no. 1, 288–292.
- [STCC12] Smith, R. E., Tournier, J.-D., Calamante, F., and Connelly, A., *Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information*, Neuroimage **62** (2012), no. 3, 1924–1938.
- [STCC13] ———, *SIFT: spherical-deconvolution informed filtering of tractograms*, Neuroimage **67** (2013), 298–312.
- [Ste65] Stejskal, E. O., *Use of spin echoes in a pulsed magnetic-field gradient to study anisotropic, restricted diffusion and flow*, The Journal of Chemical Physics **43** (1965), no. 10, 3597–3603.
- [Sug90] Sugiura, M., *Unitary Representations and Harmonic Analysis: An Introduction*, Elsevier, March 1990 (en).
- [SV01] Sethian, J. A. and Vladimirsky, A., *Ordered upwind methods for static Hamilton–Jacobi equations*, PNAS **98** (2001), no. 20, 11069–11074 (en).
- [TCC07] Tournier, J.-D., Calamante, F., and Connelly, A., *Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution*, NeuroImage **35** (2007), no. 4, 1459 – 1472.
- [TCC12] Tournier, J.-D., Calamante, F., and Connelly, A., *MRtrix: Diffusion tractography in crossing fiber regions*, Int. J. Imaging Syst. Technol. **22** (2012), no. 1, 53–66 (en).
- [TCC13] Tournier, J.-D., Calamante, F., and Connelly, A., *Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging.*, NMR in Biomedicine (2013).
- [TCGC04] Tournier, J.-D., Calamante, F., Gadian, D. G., and Connelly, A., *Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution*, NeuroImage **23** (2004), no. 3, 1176 – 1185.
- [TDV⁺14] Tax, C., Duits, R., Vilanova, A., ter Haar Romeny, B., Hofman, P., Wagner, L., Leemans, A., and Ossenblok, P., *Evaluating Contextual Processing in Diffusion MRI: Application to Optic Radiation Reconstruction for Epilepsy Surgery*, PLoS One (2014).

- [tER98] ter Elst, A. F. M. and Robinson, D. W., *Weighted Subcoercive Operators on Lie Groups*, *J Funct Anal* **157** (1998), no. 1, 88–163.
- [tHR03] ter Haar Romeny, B. M., *Front-end Vision and Multi-Scale Image Analysis: Multi-Scale Computer Vision Theory and Applications, Written in Mathematica*, Kluwer Academic Publishers, 2003.
- [TJV⁺14] Tax, C., Jeurissen, B., Vos, S., Viergever, M., and Leemans, A., *Recursive Calibration of the Fiber Response Function for Spherical Deconvolution of Diffusion MRI Data*, *NeuroImage* **86** (2014), 67–80.
- [TML11] Tournier, J.-D., Mori, S., and Leemans, A., *Diffusion tensor imaging and beyond*, *Magnetic Resonance in Medicine* **65** (2011), no. 6, 1532–1556.
- [TNRMH14] Tobisch, A., Neher, P. F., Rowe, M. C., and Maier-Hein, K. H., *Computational Diffusion MRI and Brain Connectivity*, 25–34.
- [Tsi95] Tsitsiklis, J. N., *Efficient algorithms for globally optimal trajectories*, *IEEE Transactions on Automatic Control* **40** (1995), no. 9, 1528–1538.
- [TTS⁺14] Tao, S., Trzasko, J. D., Shu, Y., Huston, J., and Bernstein, M. a., *Integrated image reconstruction and gradient nonlinearity correction.*, *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* **00** (2014).
- [Tuc04] Tuch, D. S., *Q-ball imaging*, *Magnetic Resonance in Medicine* **52** (2004), no. 6, 1358–1372.
- [TW97] Thornber, K. K. and Williams, L. R., *Characterizing the distribution of completion shapes with corners using a mixture of random processes*, *Energy Minimization Methods in Computer Vision and Pattern Recognition* (Pelillo, M. and Hancock, E. R., eds.), LNCS, no. 1223, Springer Berlin Heidelberg, May 1997, pp. 19–34 (en).
- [TYC⁺08] Tournier, J.-D., Yeh, C.-H., Calamante, F., Cho, K.-H., Connelly, A., and Lin, C.-P., *Resolving crossing fibres using constrained spherical deconvolution: Validation using diffusion-weighted imaging phantom data.*, *NeuroImage* **42** (2008), no. 2, 617–625.
- [vA07] van Almsick, M., *Context models of lines and contours*, Ph.D. thesis, Technische Universiteit Eindhoven, 2007.
- [Val14] Valkonen, T., *A primal-dual hybrid gradient method for non-linear operators with applications to MRI*, *Inverse Problems* **30** (2014), no. 5, 055012.

- [VBK13] Valkonen, T., Bredies, K., and Knoll, F., *TGV for diffusion tensors: A comparison of fidelity functions*, Journal of Inverse and Ill-Posed Problems **21** (2013), no. 3, 355–377.
- [Ves07] Veselic, K., *Spectral perturbation bounds for selfadjoint operators*, arXiv:0704.0872 [math] (2007).
- [VESB⁺13] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., and Ugurbil, K., *The WU-Minn Human Connectome Project: An overview*, NeuroImage **80** (2013), 62–79.
- [Vla06] Vladimírsky, A., *Static PDEs for time-dependent control problems*, Interfaces and Free Boundaries (2006), 281–300 (en).
- [VSJV⁺15] Van Steenkiste, G., Jeurissen, B., Veraart, J., den Dekker, A. J., Parizel, P. M., Poot, D. H. J., and Sijbers, J., *Super-resolution reconstruction of diffusion parameters from diffusion-weighted images with different slice orientations*, Magnetic Resonance in Medicine **early view** (2015).
- [WDAH13] Welsh, C. L., Dibella, E. V. R., Adluru, G., and Hsu, E. W., *Model-based reconstruction of undersampled diffusion tensor k-space data.*, Magnetic Resonance in Medicine **70** (2013), no. 2, 429–40.
- [Wei98] Weickert, J., *Anisotropic Diffusion in Image Processing*, 1998.
- [Wei99] ———, *Coherence-Enhancing Diffusion Filtering*, International Journal of Computer Vision **31** (1999), no. 2-3, 111–127 (en).
- [WHT⁺05] Wedeen, V. J., Hagmann, P., Tseng, W.-Y. I., Reese, T. G., and Weisskoff, R. M., *Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging*, Magn Reson Med **54** (2005), no. 6, 1377–1386 (eng).
- [WII99] Weickert, J., Ishikawa, S., and Imiya, A., *Linear Scale-Space has First been Proposed in Japan*, Journal of Mathematical Imaging and Vision **10** (1999), no. 3, 237–252 (en).
- [Wit83] Witkin, A. P., *Scale-space Filtering*, Proceedings of the Eighth International Joint Conference on Artificial Intelligence - Volume 2 (San Francisco, CA, USA), IJCAI'83, Morgan Kaufmann Publishers Inc., 1983, pp. 1019–1022.
- [Wol14] Wolfram Research Inc., *Mathematica*, 10.0 ed., Wolfram Research Inc., 2014.

- [XLS⁺12] Xu, J., Li, K., Smith, R. A., Waterton, J. C., Zhao, P., Chen, H., Does, M. D., Manning, H. C., and Gore, J. C., *Characterizing tumor response to chemotherapy at various length scales using temporal diffusion spectroscopy.*, PloS one **7** (2012), no. 7, e41714.
- [Yos80] Yosida, K., *Functional Analysis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1980.
- [ZDStHR16] Zhang, J., Duits, R., Sanguinetti, G., and ter Haar Romeny, B. M., *Numerical Approaches for Linear Left-invariant Diffusions on $SE(2)$, their Comparison to Exact Solutions, and their Applications in Retinal Imaging*, Numer Math Theory Me **9** (2016), no. 01, 1–50.
- [ZW04] Zweck, J. and Williams, L. R., *Euclidean Group Invariant Computation of Stochastic Completion Fields Using Shiftable-Twistable Functions*, J Math Imaging Vis **21** (2004), no. 2, 135–154 (en).

List of publications

Journal papers

J. M. Portegies, R. H. J. Fick, G. R. Sanguinetti, S. P. L. Meesters, G. Girard, and R. Duits. Improving Fiber Alignment in HARDI by Combining Contextual PDE Flow with Constrained Spherical Deconvolution. *PLoS ONE*, 10(10):e0138122, October 2015.

J. M. Portegies and R. Duits. New exact and numerical solutions of the (convection-)diffusion kernels on $SE(3)$. *Differential Geometry and its Applications*, 53:182–219, August 2017.

R. Duits, S. P. L. Meesters, J-M. Mirebeau, and J. M. Portegies. Optimal Paths for Variants of the 2D and 3D Reeds-Shepp Car with Applications in Image Analysis. *arXiv:1612.06137 [math]*, December 2016. *Accepted for publication in JMIV, Special Issue ‘Differential Geometry and Orientation Analysis’.*

E. J. Bekkers, D. Chen, and J. M. Portegies. Nilpotent Approximations of Sub-Riemannian Distances for Fast Perceptual Grouping of Blood Vessels in 2D and 3D. *arXiv:1707.02811 [math]*, July 2017. *Accepted for publication in JMIV, Special Issue ‘Differential Geometry and Orientation Analysis’.*

Conference papers

J. M. Portegies, G. R. Sanguinetti, S. P. L. Meesters, and R. Duits. New Approximation of a Scale Space Kernel on $SE(3)$ and Applications in Neuroimaging. *SSVM*, number 9087 in LNCS, pages 40–52. Springer International Publishing, May 2015.

V. Golkov, J. M. Portegies, A. Golkov, R. Duits, and D. Cremers. Holistic Image Reconstruction for Diffusion MRI. In *Computational Diffusion MRI*, pages 27–39. Springer, Cham, 2016. DOI: 10.1007/978-3-319-28588-7_3.

Abstracts

S. P. L. Meesters, G. R. Sanguinetti, E. Garyfallidis, J. M. Portegies, and R. Duits. Fast implementations of contextual PDE's for HARDI data processing in DIPY. 2016.

S. P. L. Meesters, G. R. Sanguinetti, E. Garyfallidis, J. M. Portegies, P. Ossenblok, and R. Duits. Cleaning output of tractography via fiber to bundle coherence, a new open source implementation. June 2016.

In preparation

J-M. Mirebeau, and J. M. Portegies. A fast marching solver with adaptive stencils, with applications to curvature penalized shortest paths.

Acknowledgements

Na vier jaar stug toegewerkt te hebben naar de afronding van dit boekje, kijk ik met gepaste trots op het eindresultaat dat hopelijk tot het eind der tijden een plekje krijgt in mijn boekenkast. Ik ben enorm dankbaar voor het feit dat het boekje tot stand is gekomen zonder al te veel stress, twijfels en tegenslagen waar PhDs doorgaans bekend om staan, en dat is vooral te danken aan de fijne sfeer op het werk en alle steun en aangename afleiding door familie en vrienden daarbuiten. Ik hoop dat ik de juiste woorden kan vinden om iedereen te bedanken die daartoe heeft bijgedragen, in wat misschien wel het meest gelezen hoofdstuk wordt van dit proefschrift.

Allereerst wil ik Remco bedanken, mijn dagelijkse aanspreekpunt. De duidelijke lijnen die jij had uitgezet met het project waar je mij voor aangenomen hebt, in combinatie met de enthousiaste en intensieve begeleiding, hebben ervoor gezorgd dat ik me nooit aan mijn lot overgelaten voelde. Toen ik bij jou begon als PhD, zei mijn broer: ‘ongeacht wat je bereikt met je onderzoek in de komende vier jaar, als je erin slaagt tien procent te begrijpen van waar Remco over praat heb je voldoende geleerd om tevreden terug te kunnen kijken’. Ik heb inderdaad ontzettend veel van je geleerd (ook al weet ik niet of ik dat percentage gehaald heb).

Ik ben blij dat ik, naarmate ik zelf beter mijn draai vond in het onderzoek, ook een toenemend vertrouwen van jouw kant voelde. Steeds vaker gaf je me de vrijheid om dingen op mijn eigen manier te doen, zelfs al ging dat nog regelmatig gepaard met lange discussies. Ik heb misschien nog wel de meeste bewondering voor jouw eigen drive om te blijven leren en jezelf te verbeteren. Ik hoop van harte dat je onderzoeksvoorstellen gehonoreerd worden, zodat je ons onderzoek een mooi vervolg kan geven, en ik hoop dat je de passie voor je vak nog vele jaren weet te behouden.

I would like to thank my promotor Luc Florack and all members of the committee (Kaleem Siddiqi, Giovanna Citti, Rachid Deriche, Jean-Marie Mirebeau, Barry Koren, Bart ter Haar Romeny) for taking the time to read my dissertation and for traveling (a considerable distance) to Eindhoven to be a part of the most important hour/day of my academic career. I am both grateful for and intimidated by having you as opposition.

Another thanks goes to Gonzalo. Soon after I started my PhD, you started working with Remco as a Postdoc and you took on a considerable part of supervising me. You taught me quite a bit about how to write scientifically and how to carry out experiments targeted and pragmatic. Moreover, you initiated the work on combining sub-Riemannian geometry and fast marching, that has led to what I now consider to be one of the most interesting parts of the research that I have done.

This road has also led me to Jean-Marie, without whom the results on the fast marching project would not have progressed as fast as they did. Not only did we benefit from your incredible numerical work, you also shared quite a bit of your theoretical mathematical knowledge with me, and I enjoyed very much the days I have spent in Paris to collaborate with you.

Erik, Frank, Michiel, Stefan, Stephan en Tom, de zelfbenoemde ‘Metaboyz’¹, waar ik als enige wiskundige tussen uitsluitend BMT-ers deel van uitmaakte... Bedankt dat jullie tijdens alle koppen koffie en lunches nooit te beroerd waren om mijn (gebrek aan) filmkennis belachelijk te maken, mij te helpen met mijn programmeerproblemen, lange discussies te voeren over uiteenlopende onderwerpen, en slechte grappen te vertellen.

Ik was na mijn studie waarschijnlijk niet aan de TU/e blijven plakken als mijn studentietijd niet zo leuk was geweest. Dat heb ik met name te danken aan mijn studiegenoten en vrienden, Britt, Christine, Laura en Thomas. Jullie hebben er alle vier voor gekozen om na onze studie nog door te leren en/of een PhD op te pakken, wat mij uiteindelijk heeft aangestoken hetzelfde te doen. Ik hoop dat onze vriendschap nog lang blijft, zelfs al hebben we binnenkort allemaal de TU achter ons gelaten.

Zoals de meesten die dit boekje lezen weten, ben ik buiten werktijd regelmatig op de tennisbaan te vinden. Er is geen betere uitlaatklep dan een potje tennis: de blijdschap na het slaan van een heerlijke winner en de frustratie na het slaan van een domme fout verdringen al gauw alle gedachten aan werk. Bovendien is de sfeer bij Fellenoord ongeëvenaard, waar ik alle ‘Fellies’ voor wil bedanken, en in het bijzonder mijn vaste teamgenoten Anouk, Christ, Lea, Lianne en Renze.

Ik kan me niet herinneren eerder een dankwoord gelezen te hebben waar ‘schoonfamilie’ in voorkwam. Toch hebben jullie voor mij hier zeker een plek verdiend: Hauw, Jacqueline, Mei Siu, Mei Ying, bedankt voor de interesse die jullie in mijn werk hebben getoond en alle gezelligheid in Dommelen waar ik altijd op kon rekenen.

¹naar het gebouw (MetaForum) waar we met elkaar opgescheept zaten

Opa en oma, jullie hebben iedere stap van mijn PhD op de voet gevolgd. Jullie hebben mij op het hart gedrukt dat je niet meer kan doen dan je best, wat misschien wel het beste advies is dat je een PhD kunt geven.

Jim, broer, collega, bijna-buurman. Je wiskundige kennis en begrip lijken geen grenzen te kennen, zonder dat dit ten koste gaat van je vermogen om deze kennis op anderen over te brengen. Ik weet zeker dat dit, in combinatie met je enorme toewijding, zal leiden tot een succesvolle carrière als onderzoeker en onderwijzer. Bedankt dat je mij als grote broer in zoveel dingen het goede voorbeeld hebt gegeven. Ik hoop dat je ooit de kans krijgt om voor Jason ook zo'n voorbeeld te zijn.

Marileen, lieve zus, onze PhD-tijd heeft zo'n twee jaar overlap gehad, waardoor ik bijna iedere dag op mijn werk tegen een extra 'collega' kon chatten/mopperen/grappen/... Ik liep regelmatig tegen dezelfde problemen aan als jij, en door die met jou te delen waren mijn frustraties vaak snel te relativieren. Ik hoop dat jij en Quint een mooie toekomst tegemoet gaan nadat jullie elkaar het ja-woord hebben gegeven.

Pa en ma, jullie zijn er in geslaagd om drie kinderen op de wereld te zetten die kort na het schrijven van dit boekje allemaal PhD achter hun naam mogen zetten. Hoe jullie dat voor elkaar hebben gekregen, daar zouden jullie zelf een boekje over kunnen schrijven. Met onvoorwaardelijke liefde en steun hebben jullie voor een veilige basis gezorgd waar ik altijd op kon terugvallen; een hoogtepunt deelde ik ook met anderen, maar met een tegenslag kwam ik het liefst bij jullie, en nooit stonden jullie niet voor mij klaar. Bedankt voor alles, in al die jaren.

Lieve Mei Li, natuurlijk eindig ik dit dankwoord met jou. Ik ken niemand met zoveel energie als jij. Gelukkig was jij altijd bereid om een beetje van die energie aan mij te geven, op wat voor manier dan ook. Na vier jaar onderzoek over Lie groepen ben ik er achter dat één Li voor mij genoeg is, en dat ben jij. Ik hou van je. <3

About the author

Jorg Portegies was born on 28 April 1990 in Middelburg, the Netherlands. He finished his pre-university education at the Christelijke Scholengemeenschap Walcheren in 2008, after which he started studying mathematics at the Eindhoven University of Technology. In the third year of his bachelor, he spent one semester abroad at the University of Surrey. Jorg further specialized in computational science and engineering during his master in Industrial and Applied Mathematics. His master's thesis project was a collaboration between TU/e and Philips Lighting and was focused on numerical methods for optical design. In 2013, he received his master's degree with honors.

Since 2014, Jorg has worked on his research on PDEs, group theory and applications in diffusion-weighted MRI, of which the results are presented in this dissertation. The theoretical and practical content of the research was accepted by both mathematical and biomedical journals and conferences.

Jorg will defend his PhD thesis at the Eindhoven University of Technology on 12 March 2018.

