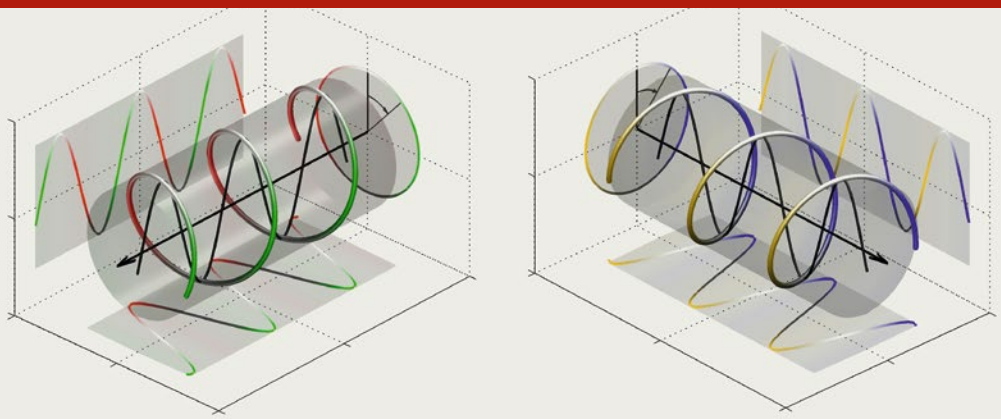
Mathematics and Visualization



Carl-Fredrik Westin · Anna Vilanova
Bernhard Burgeth *Editors*

Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data

Springer

Mathematics and Visualization

For further volumes:
http://www.springer.com/series/4562

Carl-Fredrik Westin • Anna Vilanova •
Bernhard Burgeth
Editors

Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data

With 110 Figures, 85 in color

Springer

Editors
Carl-Fredrik Westin
Dept. of Radiology
Harvard Medical School
Brigham and Women's Hospital
Boston
Massachusetts
USA

Anna Vilanova
Dept of Computer Graphics and Visualization
Technical University of Delft
The Netherlands

Bernhard Burgeth
Dept. of Mathematics
Universität des Saarlandes
Saarbrücken
Germany

ISSN 1612-3786
ISSN 2197-666X (electronic)
ISBN 978-3-642-54300-5
ISBN 978-3-642-54301-2 (eBook)
DOI 10.1007/978-3-642-54301-2
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014943920

Mathematics Subject Classification (2010): 00-XX, 20-XX, 30-XX, 32-XX, 35-XX, 51-XX, 68-XX, 92-XX

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book presents a broad and illustrative sample of the state of the art of the emerging field of visualization and processing of tensor fields and higher-order descriptors. Topics range from applications of the analysis of tensor fields to research into their mathematical and analytical properties. Tensors arise in multiple areas of mathematics, science, and technology. For example, in physics of continuous media, tensor quantities arise in constructive equations that describe charge, mass, momentum, and energy transport (the diffusion tensor, the electrical conductivity tensor, etc.). In fact, most nontrivial presentations of theories of physics resort to tensor formulations. This is true for most of the theoretical branches, such as quantum and relativistic mechanics, thermodynamics, and electromagnetism. Hence, tensor analysis is a well-established branch of mathematics that is used extensively in many areas of science and engineering. Nevertheless, tensor formulations are typically not usual in image processing and related fields, even though, for many scenarios, they should be of major interest.

The motivation is that by collecting chapters from a range of fields (e.g., including medical imaging and scientific visualization), the field of tensor processing can be more clearly presented to the interested scientific community. The field itself may be cross-fertilized with innovations bridging the various research areas. We believe that this book contributes toward this goal by combining theoretical and experimental results, reflecting recent advances in tensor signal processing and opening new avenues of research.

This book consists of the following five parts.

Part I "Tensor Data Visualization" consists of two chapters. The first chapter gives an overview of techniques for visualization of tensors and tensor fields in engineering and discusses the current state of the art and challenges. Creating visualization tools for engineering tensors often involves solving multiple different technical problems at the same time, including visual intuitiveness, interactivity, and representation of uncertainty. The second chapter is about tensor invariants and glyph design and gives an overview of common glyphs, mostly with origins in mechanical engineering, and links their interpretation to specific tensor invariants.

Part II "Representation and Processing of Higher-Order Descriptors" consists of three chapters. The first chapter describes a matrix representation of local phase, a powerful concept that has been successfully used in many image processing applications. For multidimensional signals, the concept of phase is complex and there is no consensus on the precise meaning of phase. A remedy is suggested using a novel matrix representation of multidimensional phase. In the second chapter, extension of mathematical morphological operations techniques for matrix-valued images is used and extended to be used in vector images such as color images. In the third chapter, erosion is generalized to the space of diffusion-weighted MRI data. This is done effectively by solving a Hamilton-Jacobi-Bellman system (erosion) on the coupled space of three-dimensional positions and orientations, embedded as a quotient in the group of three-dimensional rigid body motions. The solution to the HJB equations is given by a well-posed morphological convolution.

Part III "Higher-Order Tensors and Riemannian-Finsler Geometry" includes four chapters that provide powerful mathematical language to model and analyze large and complex diffusion data such as high angular resolution diffusion imaging (HARDI) or diffusion kurtosis imaging (DKI). The first chapter gives a careful introduction to the foundations of higher-order tensor algebra and explains how some concepts from linear algebra generalize to the higher-order case. The second chapter discusses fourth-order symmetric tensors and how to model the positivity constraint present in diffusion. The third chapter describes the role of Riemann-Finsler geometry as a potentially powerful mathematical framework in the context of diffusion MRI and presents the basic theoretical foundation for Finsler-based tractography. The fourth chapter expands on these concepts and presents more details on a numerical Finsler tractography implementation.

Part IV "Tensor Signal Processing" presents new methods for processing of tensor-valued data. The first chapter gives a novel perspective on performing voxel-wise morphometry of diffusion tensor data using kernel-based approach. The second chapter reviews the free-water diffusion model and uses it to derive diffusion tensors following the elimination of the free-water component that is assumed to originate from the extracellular space. The third chapter reviews approaches that have been proposed to compute fabric tensors with emphasis on trabecular bone research. Fabric tensors aim to model both anisotropy and orientation of a material with respect to another one.

Part V "Applications of Tensor Processing" illustrates the use of tensors in different application domains. The first chapter reviews various types of tensors used in geometry processing and discusses the use of the metric and curvature tensors, two of the most studied tensors in geometry processing. The final two chapters demonstrate applications of diffusion-weighted imaging toward clinical use. One chapter describes a probabilistic technique for diagnostic prediction of first-episode schizophrenia patients based on their brain diffusion MRI data. The last chapter presents an interactive system that integrates the visual analysis of diffusion MRI tractography with data from electroencephalography (EEG).

We believe this collection of chapters captures the excitement and inspiration that has been generated during a series of Dagstuhl seminars devoted to visualization and

processing of tensor fields. This book includes contributions from attendees of the fourth meeting, entitled "Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data," held in December 2011. As in the three earlier volumes, the authors report on recent research results and future directions for the analysis and visualization of tensor fields. One of the goals of this seminar is to bring together researchers from along the axis between pure and applied research, identifying new multidisciplinary research challenges. This book, we hope, will continue to further that goal in a broader context.

Boston, USA — Carl-Fredrik Westin
Eindhoven, The Netherlands — Anna Vilanova
Saarbrücken, Germany — Bernhard Burgeth
October, 2013

Contents

Part I
Tensor Data Visualization

Top Challenges in the Visualization of Engineering Tensor Fields

Mario Hlawitschka, Ingrid Hotz, Andrea Kratz, G. Elisabeta Marai, Rodrigo Moreno, Gerik Scheuermann, Markus Stommel, Alexander Wiebel, and Eugene Zhang

Abstract In this chapter we summarize the top research challenges in creating successful visualization tools for tensor fields in engineering. The analysis is based on our collective experiences and on discussions with both domain experts and

M. Hlawitschka (✉) • G. Scheuermann
Department of Computer Science, Leipzig University, Leipzig, Germany
e-mail: hlawitschka@informatik.uni-leipzig.de; scheuermann@informatik.uni-leipzig.de

I. Hotz
German Aerospace Center, Braunschweig
e-mail: ingrid.hotz@dlr.de

A. Kratz
Zuse Institute Berlin, Berlin, Germany
e-mail: kratz@zib.de

G.E. Marai
Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15260, USA
e-mail: g.elisabeta.marai@gmail.com

R. Moreno
Department of Medical and Health Sciences (IMH) and Center for Medical Image Science and Visualization (CMIV), Linköping University, Campus US, 58185 Linköping, Sweden
e-mail: rodrigo.moreno@liu.se

M. Stommel
Chair of Polymer Materials, Saarland University, Saarbruecken, Germany
e-mail: m.stommel@mx.uni-sarland.de

A. Wiebel
Department of Electrical Engineering and Computer Sciences, Coburg University of Applied Sciences, Coburg, Germany
e-mail: alexander.wiebel@hs-coburg.de

E. Zhang
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97330, USA
e-mail: mzhange@eecs.oregonstate.edu

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_1,

visualization practitioners. We find that creating visualization tools for engineering tensors often involves solving multiple different technical problems at the same time—including visual intuitiveness, scalability, interactivity, providing both detail and context, integration with modeling and simulation, representing uncertainty and managing multi-fields; as well as overcoming terminology barriers and advancing research in the mathematical aspects of tensor field processing. We further note the need for tools and data repositories to encourage faster advances in the field. Our interest in creating and proposing this list is to initiate a discussion about important research issues within the visualization of engineering tensor fields.

1 Introduction

Many engineering disciplines make extensive use of tensors [2]. For example, applications in solid mechanics, civil engineering, bioimaging and bio engineering, computational fluid dynamics, geology and electrical engineering require the processing of tensor fields as part of domain-specific modeling, simulation, and analysis processes.

Given the complexities of engineering tensor data—including large-scale—, visualization can be a powerful ally. In the last few years, this fact has fostered research in the visualization and processing of engineering tensor fields within the visualization community. The importance of these efforts relies on the huge potential impact of using advanced techniques of visualization in helping engineering professionals and scientists to have a better understanding of problems that involve processing of tensor fields. In addition, including features such as interactivity in the processing pipeline can be used to improve the results yielded by numerical simulations.

Visualization of engineering tensor fields is, however, a relatively new research topic [3, 5, 11]. Despite the potential advantages of tensor visualization in engineering, significant challenges make advances in the field difficult. Asking ourselves what the most important research challenges facing us are, and identifying the stumbling blocks, as well as the required practices, has the potential to speed up our progress. In this chapter, we attempt to start a discussion of these issues by proposing a list of top research challenges and issues in the visualization of engineering tensor fields.

We have been assembling this list of challenges over a series of discussions which included both visualization researchers and domain practitioners. Our interest in creating and proposing this list is not to impose our own ideas on the field, but rather to jump-start a discussion about important research issues [6] within the visualization of engineering tensor fields. We expect that this list will grow as the field itself grows and as additional topics are identified.

Here follows our list of the top challenges in the Visualization and Processing of Engineering Tensors. Please note that this list is not ranked.

2 Mathematical Models

Tensors are perhaps one of the most commonly used concepts in physics, geometry, engineering, and medical research. There has been much research in tensors and tensor fields in terms of their mathematical and numerical analysis as well as geometric and physical properties. In the visualization community much groundwork has been carried out in the visualization of tensor fields. However, there are significant gaps between what we already know about tensors and what we can do with them, both in the visualization community and the application domains. This is due to a number of challenges that we face today.

First, mathematical analysis and visualization of tensor fields is intrinsically difficult, thanks to the large amount of information contained even in a single second-order tensor (four numbers in 2D and nine in 3D). Existing analysis mainly focuses on local (e.g., pointwise) tensor properties, and relatively little work exists in global tensor field analysis, such as tensor field topology. Even in low-dimensional cases, typical tensor data can consist of millions of nodes, each of which is associated with a tensor. The sheer amount of data poses great challenges in robust analysis and effective visualization of tensor fields.

Second, mathematical analysis of tensor fields is further complicated by the order of tensors. The order of a tensor refers to the number of indices that is needed to describe the entries in the tensor. Zeroth-, first-, and second-order tensors refer to scalars, vectors, and matrices. While there is great need to handle tensors of order higher than two, our ability to process such tensors are rather limited. For example, spectral analysis for higher-order tensors is more difficult than that for second-order tensors, and it has not been found how to extend the notion of eigenvalues and eigenvectors to higher-order tensors that would satisfy all the properties that eigenvalues and eigenvectors of second-order tensor fields possess. Moreover, global structures in higher-order tensors are not well understood, and it is not yet clear what information is essential for understanding higher-order tensors. In addition, higher-order tensors contain more entries, which makes the storage and effective processing more challenging than their lower-order counterparts such as vectors and matrices.

3 Relevant Quantities and Terminology

In many application areas, e.g., solid mechanics, the use of tensor related quantities has a long tradition. In these areas a very application specific terminology has been developed. Thus often a multiple of identifiers or names exist for the same mathematical entity. This is already the case for basic entities such as eigenvalues: in the context of stress tensors in engineering eigenvalues are usually called principal stresses. Even more confusing, there are terms used across disciplines which have different definitions, e.g. anisotropy.

A good visualization is guided by the practical questions and relevant quantities of the application. Even more than in other visualization areas, the questions related

to tensor fields strongly vary across application fields. Similar to the terminology, the relevant quantities cannot easily be transferred. The physical meaning of the mathematical properties deviates strongly. For example, while in some applications isotropic or degenerate points (points where the eigenvectors are not uniquely defined) are of special importance, in other contexts they are just points of high symmetry without particular meaning. We note that the terminology of tensor fields in mathematics, the physical (or application-dependent) interpretation of tensor fields in application areas, and the terminology from the application areas are all correct, from a reasonable perspective. However, clear translations of the various definitions across fields are needed.

In summary, it is difficult to transfer visualization methods from one application to another without specific adaptation. Entering a new application area always requires significant effort to get to know the domain-specific language, which can be discouraging to visualization researchers. Finally, accessing the relevant information is often not trivial since it is scattered in the literature and rarely do any concise introductions [1, 8] exist.

4 Effective Visual Abstractions

Effective visual abstractions are a fundamental problem in the visualization of engineering tensors. By *effective* we denote visual abstractions which capture the physical or mathematical aspects of the tensors and which are intuitive to the application-domain practitioners.

The visual abstraction issue is particularly difficult because the physical meaning of engineering tensors is not necessarily intuitive. Symmetric second order tensors are used routinely as abstract quantities in the mathematical modeling of turbulent combustion and considered very useful for computation. Yet only abstractions of the tensor, such as the trace of the tensor matrix, may bear physical meaning to the domain practitioner. In this context, it is important to be able to select from the many visual abstractions that have already been proposed, and to effectively combine them into frameworks that solve specific engineering problems.

The physical meaning of tensors can further greatly impact how they should be analyzed and visualized, even when the mathematical representations of these tensors are the same. Examples of this include the stress tensor and strain tensor from solid mechanics, the rate of deformation tensor from fluid dynamics, and the diffusion tensor from medical imaging, all of which are second-order, symmetric tensors. Yet, different mathematical analyses and visualizations are needed that best suit the domain scientists' needs. While certain tensor visual abstractions have been established in other application domains, they do not necessarily transfer well to engineering. For example, unlike diffusion tensors in medical imaging, the eigenvectors and eigenvalues of engineering tensors may be meaningless in turbulent combustion; while reduction of the tensor field to a scalar field—e.g., stress in mechanical engineering, or divergence in turbulent combustion—may bear particularly intuitive meaning to the engineering practitioner. To create effective

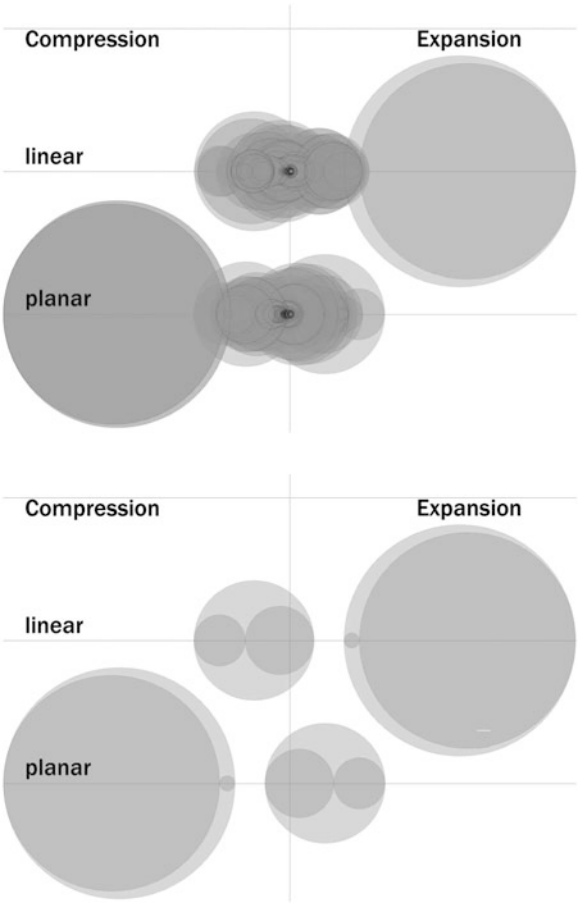


Fig. 1 Mohr diagram of a two-force dataset (simulation of a block on which a pushing and a pulling force are applied); *top*: unfiltered Mohr diagram, *bottom*: filtered Mohr diagram. The filtered Mohr diagram depicts only *circles* for cluster representatives that have been computed using mean-shift clustering in shape space. This representation reveals that four characteristic stresses occur: Compressive, tensile and mixed stresses. These characteristics are found in many engineering datasets

visual representations, visualization researchers need to spend significant time understanding the underlying science and engineering.

Finally, the particular visual abstraction depends on which aspects of the model are considered important. Finding effective, understandable visual abstractions for engineering tensors is thus a fundamentally interdisciplinary, exploratory process. Domain hypotheses and data change iteratively with the visual exploration process, thus the data models and thus visual abstractions change iteratively as well. Communication with the domain experts is needed to establish which aspects of the model are important and need to be captured by the visual representations.

To identify the right or interesting quantities to visualize and to establish a common visual language with the domain specialists, visualization researchers have found it useful to cover the domain literature and seek traditional visual representations (Fig. 1) [7]. Another approach is to tap into information visualization abstractions such as icons or glyphs through a parallel prototyping [4] approach. However, identifying the right invariants and visualization rule-of-thumbs for a particular domain—and then across engineering fields—remains a major challenge.

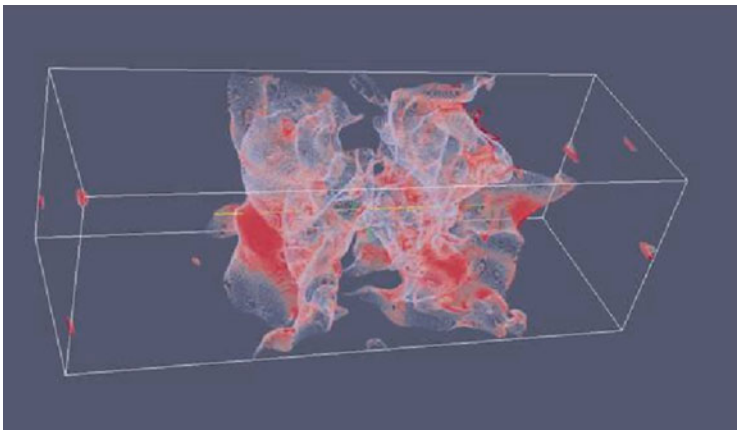


Fig. 2 Paraview rendering of the shock regions (points within 0.01 of the local speed/speed of sound ratio of 1) in a rich and very large—21M-points—combustion dataset. Due to the supersonic nature of this test case, the flow field exhibits shocklets during the simulation

5 Scalability: Very Large Datasets

Engineering tensor data tend to be generated from computational simulations. Major advances in computing capabilities mean that recent datasets tend to have very large-scale—petascale and exascale: even toy examples from turbulent combustion have millions of grid points (Fig. 2).

Given the scale of these tensor fields, major scalability challenges include occlusion and clutter. For example, information encoded using glyph-representations becomes quickly unreadable. Furthermore, slow interaction with complex, though expressive geometric representations means hardware-accelerated techniques are needed to render and explore the data at interactive rates.

Possible solutions to scalability issues include abstraction (volume rendering [7, 9], Fig. 3), interactive filtering (also projection to planes), clustering (also dimension reduction), or simply creating novel scalable visual representations. It may also be useful to borrow scalable ideas from machine learning and information visualization.

6 Scalability: Multi-field Datasets

In addition to large-scale, engineering data also typically feature multiple fields, including non-tensor quantities such as pressure, temperature, or velocity. For example, mechanical engineering problems may feature more than 60 tensor and non- tensor field quantities in the same dataset [10]. Furthermore, numerical simulations may run over many time steps, in which case comparison of tensor

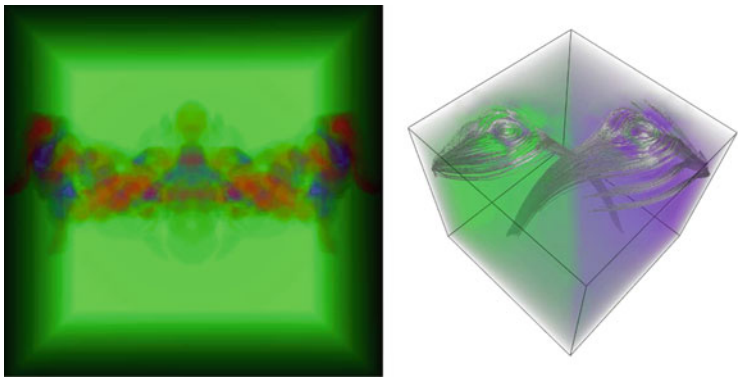


Fig. 3 *Left*: Volume rendering of a three tensor clustering for a 8M point mixing-layer combustion dataset (clusters mapped to *green*, *red*, and *blue*). *Right*: Hybrid Rendering of the two-force dataset. Volume rendering provides context (*violet* = tension, *green* = compression) and tensor lines are seeded for the pushing and pulling force only (along major and minor eigenvector, respectively)

fields across time becomes particularly important. We should be able to effectively visualize multiple fields simultaneously, and to visualize the interaction between multiple fields.

However, abstractions which meet successfully the large data scalability challenge—such as transfer-function based volume rendering or hybrid representations—do not necessarily facilitate the visual comparison of multiple fields at the same time. Established techniques for visual comparison—such as juxtaposition, overlays or animation—may require novel visual abstractions to meet the challenges of tensor multi-field data that stem from engineering fields.

7 Scalability: Details in Context

The shear spatial size combined with many small-scale phenomena provides additional challenges: Whereas details-in-context techniques have a long history in medical visualization and in information visualization, a transfer of those techniques to engineering data is typically not a trivial task that is made even harder due to the lack of dataset independent spatial reference other than the provided geometry.

The delivery of visualization systems that follow basic interaction paradigms such as providing focus and context in tensor fields at the same time often rely on the possibility to display data at different levels of detail, and pan and zoom techniques require relatively smooth transitions between representations at different scales. Besides modified seeding strategies for glyphs and topology simplification, only few techniques currently used in tensor visualization fulfill these requirements and, to the best of our knowledge, none of these has been studied in detail in this context.

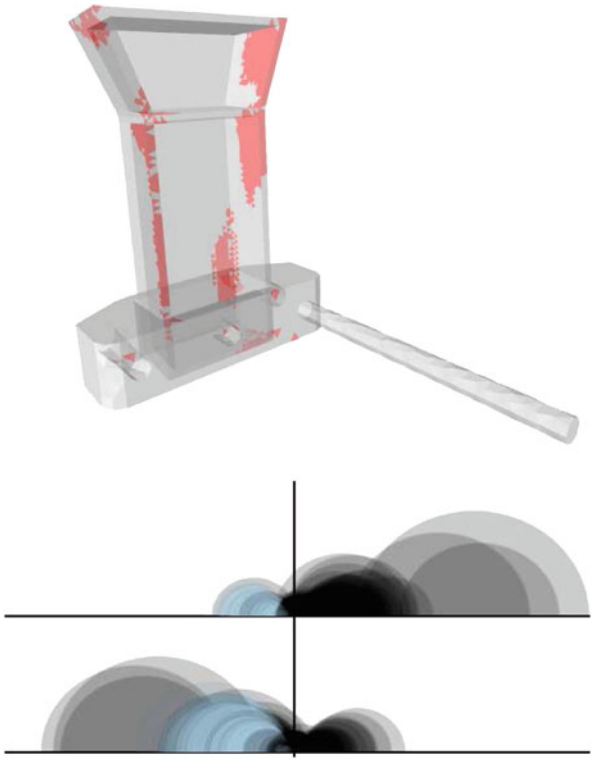


Fig. 4 Construction element dataset used for modeling and simulation in mechanical engineering. Visualization should be integrated with the modeling and simulation process, and not a post-processing step. *Top*: Compressive areas are highlighted. *Bottom*: Mohr diagram. The dataset exhibits the same basic characteristics as the two-force dataset

8 Integration of Visualization with Simulation and Modeling

Most engineering problems do not require mere post-processing visualization, but interactive visual computing (Fig. 4). For example, tensor visualization is fundamentally interesting for the validation or debugging of numerical simulations in the context of computational turbulent combustion modeling. In such situations, the domain experts are often interested in visually exploring tensor datasets as they are being generated by the numerical method used in the simulation. Detecting anomalies in the tensor field may highlight bugs in the modeling stage or in the numerical simulation stage. Early detection of such anomalies may help stopping in such cases lengthy, computationally expensive simulations that would lead to erroneous final results.

With simulation runtimes on the order of days or months even on supercomputers, the integration of visual analysis with the simulation and modeling processes could lead to significant benefits. Furthermore, the domain experts are interested in steering calculations in real-time, being able to change parameters on the fly and see the effects.

However, the scale of these simulations and the hardware on which they are run pose stringent constraints. Massively parallel simulations are common: How do we

combine the results from such parallel simulations in an interactive rendering of the entire dataset? Furthermore, large-scale datasets often cannot be saved to disk: How do we visualize tensor data when the domain expert cannot afford to stop the simulation to save the data for visualization? In situ visualization and remote visualization of tensor fields are, in this context, topics of high interest.

9 Interpolation and Smoothing

tensor field data can come from direct measurement or numerical simulation. In both cases the data is only available at discrete locations (e.g., the vertices of a grid or integration points of a cell). However, most tensor field analysis and visualization approaches assume an everywhere continuous tensor field. Consequently, it is necessary to reconstruct the tensor data from the vertices to the rest of the domain (e.g., points on the edges, faces, and interiors of grid cells). This process, referred to as interpolation or extrapolation, requires great care. While it seems simple and is rather straightforward to perform numerically, factors such as the size of the grid and its configuration as well as the used interpolation scheme are often ignored even though they impact the amount of faithfulness of the interpolated data with respect to the ground truth. However, without this being properly understood, errors may be introduced in the data that compromise the quality of the subsequent physical interpretation. We wish to point out that the difference between the ground truth and the interpolated tensor fields is difficult to quantify. It is similar to object reconstruction from photos, where the ground truth is generally not known, and therefore the difference between it and the reconstructed object is not available. Moreover, it is not clear how to reduce this difference. Consequently, the difference can be viewed as a form of error, or, as we argue here, a form of uncertainty.

Another useful and often necessary data processing operation is tensor field smoothing. This operation is designed to remove high-frequency signals from the data, under the assumption that such signals are noise and therefore of little relevance to the physical interpretation. However, such interpretation has not been validated for tensor field smoothing by the visualization community. It is not clear how the global structures (topological) in the tensor fields are impacted by the tensor field smoothing. In addition, there has not been a principled way of deciding how much smoothing is needed. Consequently, uncertainty is introduced into the data.

Another challenge is which mathematical representation for tensors should be used during smoothing. A popular approach is to perform tensor field smoothing on the entries of the tensors. While this seems to work well in practice in many applications, it is nonetheless often unjustified. Other approaches, such as smoothing on the norm, the eigenvalues and the eigenvectors of the tensor, can and should be explored and compared with smoothing on the tensor entries.

10 Uncertain Data

The fact that data is often given in a discrete form and the resulting need for interpolation is only one source of error found in simulated and measured data. As neither simulations nor measurements can be taken as ground truth, the introduced error can be handled as one kind of uncertainty in the data. In addition, a variety of errors accumulate in a data processing pipeline, most of which can be described roughly at the place of occurrence: i.e., the distortion of the measurement, measurement bias, discretization and quantification errors, calculation errors in pre-processing or simulation due to the limited mathematical, and, last but not least, the graphical precision of the screen.

Data processing pipelines, including visualization systems, would greatly profit from propagating errors from the step where they occur to the visualization and incorporating these errors in the visualization; whereas to date, the most-common approach is to keep every single error small and then ignore them. Error bars are the most common representation in all scientific journals but equivalent metaphors for the visualization of tensor data are rare.

The uncertainty of input data has further implications to the processing when combined with interpolation. Where interpolation in certain data leads to errors (or uncertainties) between the sample points, interpolation (seen as averaging) of uncertain data may actually reduce uncertainty. Whereas such models are being explored for scalar data, an application to tensor-valued data remains an open question and one of the most important challenges for the next years.

11 Data Repositories

Progress in many scientific disciplines benefits from the availability of free data or benchmarks. This is especially true for visualization where data is the starting point for all techniques. The impact of such data is threefold: The data can be used to develop and test new techniques. It can help to show new research directions and demonstrate shortcomings of existing methods. Finally, it can be used to evaluate the applicability of new techniques, as well as their accuracy and efficiency compared to existing ones.

Looking at visualization research, freely available data did accelerate the developments at least in volume visualization (http://www.volvis.org, and others) and flow visualization (Vis-Contests 2004, 2006, 2011, and the International CFD Database at http://cfd.cineca.it/). Although not in one repository like volvis.org, for tensor visualization there exist a large number of freely available data sets from the medical domain (e.g. Kindlmann's data page, http://www.grand-challenge.org and Vis-Contest 2010) and some datasets related to earthquakes. However, scanning through the tensor visualization literature shows that there are only very few and, maybe more importantly, very simple datasets from engineering applications

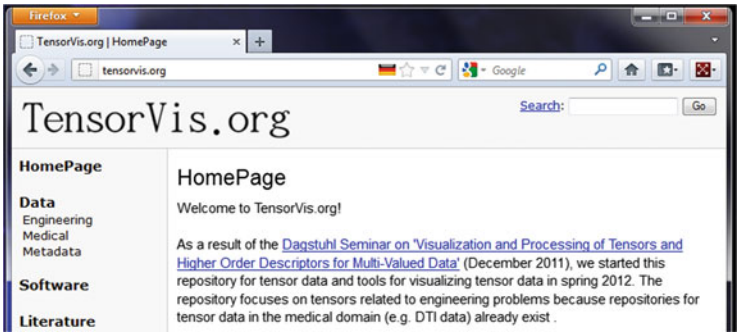


Fig. 5 Screenshot of the tensorvis.org website that hosts the prospective data and tool repositories

available. As the notion and use of tensors is very different in different application domains, this is a severe problem for the development of visualization techniques for such data. As a first step to mitigate this problem, we set up a data repository at http://www.tensorvis.org (Fig. 5) in Spring 2012 and work towards populating it with various tensor datasets from engineering applications and meta information about these datasets. Maintaining minimum meta information requirements we hope to make the provided data even more useful.

12 Tool Repositories

In addition to the availability of data, research can also be supported and even accelerated by the availability of software tools. Again, the impact is three-fold: Tools can help to convert, load and process the data to be examined, tools providing visualizations themselves can be used as benchmarks for new techniques, and, finally, researchers can integrate some tools or libraries into their own software.

Domain practitioners new to the field of tensor visualization are often unaware of the various existing tensor visualization tools and tensor processing libraries. A repository providing descriptions and web-links for download or further information would make their first steps in the field easier. Such repositories should be integrated with data repositories to have all relevant information in one place.

13 Conclusion

Analyzing the collection of challenges we introduced in this chapter, we note that creating visualization tools for engineering tensors often involves solving multiple different technical problems at the same time—including visual intuitiveness, scalability, interactivity, providing both detail and context, integration with modeling

and simulation, representing uncertainty and managing multi-fields. This level of complexity usually leads to trade-offs among different strategies that aim at reasonably tackling all technical problems in the same framework. In addition, many mathematical aspects of tensor field processing, which are also necessary for visualization, are still under active research.

We further note that the gap between the availability of visualization tools and their actual use by domain practitioners on a regular basis is still huge. Although designing application-dependent visual abstractions is one of the strategies that can help to reduce such a gap, this usually requires visualization researchers to have a very good knowledge of the specific application, which is time consuming and not always effective. Finally, as in many other research areas, the lack of tools for sharing knowledge within the community has hitherto discouraged faster advances in the field.

In this chapter, we aimed to summarize the challenges in creating successful visualization tools for tensor fields in engineering. With this analysis and formalization of our collective experiences in the visualization of engineering tensors, we hope to motivate visualization researchers to think either about new tensor-related problems or about persistent tensor problems across engineering fields with a refreshed perspective.

Acknowledgements Grateful acknowledgments to H. Hagen (Kaiserslautern University), S. Levent Yilmaz, Mehdi Nik, Tim Luciani, Adrian Maries and Md. Abedul Haque (Pitt) for gracefully providing several of the images and captions in this chapter, as well as for inspiring discussions. G.E. Marai's work is partially supported through NSF IIS-0952720 and NSF CBET-1250171.

References

1. Brannon, R.: Mohr's circle and more circles (Online). Available: http://www.mech.utah.edu/brannon/public/MohrsCircle.pdf (2003)
2. Delmarcelle, T., Hesselink, L.: The topology of symmetric, second-order tensor fields. In: IEEE Visualization 1994, Washington, DC, pp. 140–147 (1994)
3. Dick, C., Georgii, J., Burgkart, R., Westermann, R.: Stress tensor field visualization for implant planning in orthopedics. IEEE Trans. Vis. Comput. Graph. **15**(6), 1399–1406 (2009)
4. Dow, S., Glassco, A., Kass, J., Schwarz, M., Schwarz, D.L., Klemmer, S.R.: Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. Trans. Comput. Hum. Interact **11**(4), 18:1–18:24 (2010)
5. Hashash, Y.M.A., Yao, J.I.-C., Wotring, D.C.: Glyph and hyperstreamline representation of stress and strain tensors and material constitutive response. Int J Numer Anal Methods Geomech **27**(7), 603–626 (2003)
6. Johnson, C.R.: Top scientific visualization research problems. IEEE Comput. Graph. Appl. **24**(4), 13–17 (2004)
7. Kratz, A., Meyer, B., Hotz, I.: A visual approach to analysis of stress tensor fields. ZIB-report technical report 10–26 (2010)
8. Kolecki, J.C.: An introduction to tensors for students of physics and engineering, NASA/TM–2002-211716 (Online). Available: http://www.grc.nasa.gov/WWW/k-12/Numbers/Math/documents/Tensors_TM2002211716.pdf

9. Maries, A., Haque, Md.A., Yilmaz, S.L., Nik, M.B., Marai, G.E.: Interactive exploration of stress tensors used in computational turbulent combustion. In: Laidlaw, D., Villanova, A. (eds.) New Developments in the Visualization and Processing of Tensor Fields. Springer, Heidelberg (2011)
10. N.N.: Abaqus Analysis User's Manual, Version 6.7 EF. Dassault Systèmes, Simulia Corp., Providence
11. Slavin, V., Pelcovits, R., Loriot, G., Callan-Jones, A., Laidlaw, D.: Techniques for the visualization of topological defect behavior in nematic liquid crystals. IEEE Trans. Vis. Comput. Graph. **12**(5), 1323–1328 (2006)

Tensor Invariants and Glyph Design

Andrea Kratz, Cornelia Auer, and Ingrid Hotz

Abstract Tensors provide a mathematical language for the description of many physical phenomena. They appear everywhere where the dependence of multiple vector fields is approximated as linear. Due to this generality they occur in various application areas, either as result or intermediate product of simulations. As different as these applications, is the physical meaning and relevance of particular mathematical properties. In this context, domain specific tensor invariants that describe the entities of interest play a crucial role. Due to their importance, we propose to build any tensor visualization upon a set of carefully chosen tensor invariants. In this chapter we focus on glyph-based representations, which still belong to the most frequently used tensor visualization methods. For the effectiveness of such visualizations the right choice of glyphs is essential. This chapter summarizes some common glyphs, mostly with origin in mechanical engineering, and link their interpretation to specific tensor invariants.

1 Introduction

As multilinear mappings, tensors are an important tool in many physical and engineering disciplines. They appear everywhere, where a linear correspondence between vectorial entities is assumed. Thus, they play a large role for many simulations of physical phenomena either as the result or as an intermediate product. For an introduction to and an overview over recent work concerning the

A. Kratz (✉) • C. Auer
Zuse Institute Berlin, Berlin, Germany
e-mail: kratz@zib.de; auer@zib.de

I. Hotz
German Aerospace Center, Braunschweig, Germany
e-mail: ingrid.hotz@dlr.de

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_2,

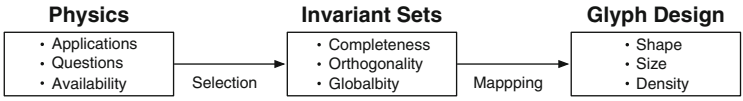


Fig. 1 Based on applications and their questions, meaningful invariant sets are extracted. Then selected sets are mapped to geometric shapes, which depict the tensor. Mathematically, the most efficient sets are complete and orthogonal. For the design of glyphs, however, other sets can be preferable

visualization of tensor fields we refer to [20]. Concerning the analysis of tensors, practical questions are, in general, related to a set of specific tensor invariants that describe key aspects of the application. Visualizations that build on these sets of invariants are close to the everyday language of engineers and thus capable of supporting the data analysis in an intuitive way. Central to our work is the finding that tensor visualization methods can be efficiently designed and parametrized by an appropriate choice of invariants.

Due to its high information content, it is difficult to visualize the whole tensor information in a continuous way. Therefore tensor field visualization is always a compromise between a continuous, global representation of partial information or a discrete, local representation of the entire information. Local methods depict single tensors sampled at discrete positions across the field, where geometric objects (glyphs) are used to encode tensor properties in terms of shape, color and orientation. In this chapter we focus on such local glyph-based visualizations. In order to provide easy-to-read glyphs, it is crucial to decide, *what* information to choose, *how* to map the information onto glyph parameters [23] and *where* to place the glyph (see also Fig. 1). The goal of this chapter is to give an overview of commonly used glyphs in different mechanics applications and put them into context with relevant tensor invariants whenever possible. We also include some important sets of invariants, for which there are no appropriate glyphs encoding them directly yet.

2 Basics and Notations

We restrict our considerations to second-order tensors given in three-dimensional space. A second-order tensor $\mathbf{T}$ is defined as a bilinear function from two copies of a vector space V into the space of real numbers, $\mathbf{T} : V \otimes V \to \mathbb{R}$. Another but equivalent perspective defines a tensor $\mathbf{T}$ as linear operator that maps every vector onto a vector, i.e. a map from vector space V onto itself, $\mathbf{T} : V \to V$. Using a specific basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ of the vector space V, $\mathbf{T}$ can be represented as matrix. A tensor is uniquely determined by its 3 *eigenvalues* (*principal values*) λ_i and corresponding *eigenvectors* (*principal directions*) $\mathbf{e}_i$. The eigenvalues are defined as the roots of the characteristic polynomial $\det(\mathbf{T} - \lambda\mathbf{I}) = 0$.

Considering different application areas, tensors exhibit specific properties that have an impact on the choice of an appropriate visualization method. In the following, we summarize the most important tensor properties.

Symmetry A tensor is called symmetric if its matrix representation is symmetric, this is $m_{ij} = m_{ji}$ for all i, j. For symmetric tensors the eigenvalues are real and the eigenvectors mutually orthogonal. With respect to the basis spanned by the eigenvectors, the tensor matrix is diagonal. Physically, the eigenvectors describe the directions of extremal variations of the quantity that is encoded by the tensor. The sign of the eigenvalues often has a distinguished physical meaning.

Definiteness We distinguish between positive (semi-) definite, negative (semi-) definite and indefinite tensors. A tensor is *positive definite* if the characteristic invariants of $\mathbf{T}$ are all positive, i.e. all eigenvalues of $\mathbf{T}$ are positive. It is *positive semi-definite* if all eigenvalues are larger than or equal to zero. The definition for *negative (semi-) definite* tensors is analogue. If the tensor's eigenvalues are positive as well as negative, it is called *indefinite*.

Traceless Considering the tensor $\mathbf{T}$, represented by a matrix $M \in \mathbb{R}^{3\times3}$, the trace is defined as the sum of its diagonal components $tr(M) = \sum_{i=1}^{3} m_{ii}$. Tensors with trace zero are called *traceless*.

Decompositions A decomposition into an *isotropic* and an *anisotropic part* is useful for many applications. Thereby, one has to distinguish the different classes of tensors: Positive-definite tensors with a character similar to deformation separate multiplicatively. This decomposition is also called *dilation-distortion* or *volumetric-isochoric* $\mathbf{T} = \mathbf{T}^{dil} \cdot \mathbf{T}^{dist}$. The dilatation part is related to volume changes. It is given as $\mathbf{T}^{dil} = J^{1/3}\mathbf{I}$, where $J = \det(\mathbf{T})$. The distortional or isochoric part accounts for shape changes given by $\mathbf{T}^{dist} = J^{-1/3}\mathbf{T}$. Indefinite tensors as, e.g., natural strain or stress, with a deformation generating character, separate additively $\mathbf{T} = \mathbf{T}^{dev} + \mathbf{T}^{iso}$. Where $\mathbf{T}^{dev} = \mathbf{T} - \frac{1}{3}\mathrm{tr}(\mathbf{T})\mathbf{I}$ is called *anisotropic part* or *deviator* and $\mathbf{T}^{iso} = \frac{1}{3}\mathrm{tr}(\mathbf{T})\,\mathbf{I}$ *isotropic* part. From a physical point of view, the isotropic part $\mathbf{T}^{iso}$ represents a direction-independent transformation (e.g., a uniform scaling or uniform pressure) and the deviatoric part $\mathbf{T}^{dev}$ represents the distortion. In context of stress, the deviatoric part is often analyzed to identify material failure.

3 Tensors in Mechanical Engineering

Among the most commonly used tensors of second order in mechanical engineering one can roughly distinguish three different classes. The first class are tensors that describe a change of a state, e.g., related to deformations; the second class are tensors generating a change of state, related to forces; the third class approximates distributions, e.g. of orientations. In the following we introduce some of the most relevant tensors in their physical context. A comprehensive introduction to this topic can be found in [6].

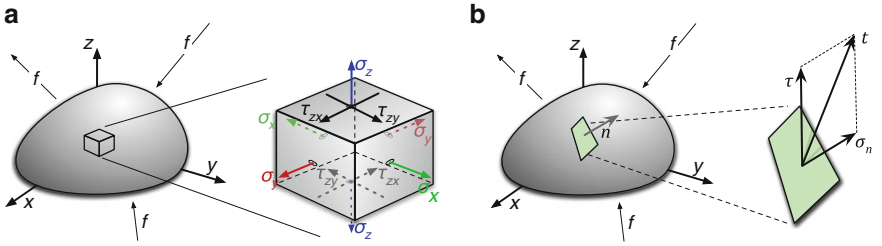


Fig. 2 Stress tensor. External forces f are applied to a deformable body (**a**). The reacting forces are described by a three-dimensional stress tensor. The tensor is composed of three normal stresses σ and three shear stresses τ. (**b**) Given a surface normal n of some cutting plane, the stress tensor maps n to the traction vector t, which describes the forces that act on this plane (normal and shear forces)

Deformation gradient tensor Continuum mechanics deals with the analysis of deformable bodies. The deformation is described by the deformation gradient tensor $\mathbf{F}$. It is defined as the gradient of displacements of material points. Since no cell inversions are allowed inside the material, the tensor is invertible and has a positive determinant. It quantifies shape changes and material rotation and, thus, is not symmetric. If (x_1, x_2, x_3) are the original coordinates and (X_1, X_2, X_3) the corresponding coordinates after deformation it is defined as $\mathbf{F} = \left(\partial X_i / \partial x_j\right)$.

Strain tensor The strain tensor can be derived from the deformation gradient tensor. Compared to the deformation gradient tensor it loses the information about rotation. It expresses the relative length and orientation changes during deformation. It is symmetric and indefinite. There are different ways to define strain explicitly. A common definition with respect to the original reference frame is given by $\mathbf{E} = \frac{1}{k}((\mathbf{F} \cdot \mathbf{F}^T)^k - \mathbf{I})$ with $k \in N$. $\mathbf{F}^T$ is the matrix transposed to $\mathbf{F}$ and "$\cdot$" the standard matrix multiplication. For $k = 2$ this is the Lagrangian strain, for $k \to 0$ it becomes the natural strain. All of these strain definitions are equivalent if the deformations are infinitesimal small.

Stress tensors The stress tensor describes internal forces or stresses that act on a material within a deformable body as reaction to external forces (Fig. 2a). If forces are balanced and there is no rotation (which is, in general, fulfilled for infinitesimally small volume elements), the stress tensor is symmetric. The sign of the eigenvalues differentiates compressive or tensile forces, whereby, there is no unique sign convention. Stress has the dimension of pressure, its unit is force per area. Considering cutting planes with normal $\mathbf{n}$ through the material, the forces acting on these planes are given by the traction vector $\mathbf{t}$ (Fig. 2b).

$$\mathbf{t} = \mathbf{T} \cdot \mathbf{n} = \begin{pmatrix} \sigma_{xx} & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \sigma_{yy} & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \sigma_{zz} \end{pmatrix} \cdot \mathbf{n} = \tau_n + \sigma_n. \tag{1}$$

The traction vector can be decomposed into a part normal to the cutting plane $\sigma_n = (\mathbf{n}^T \cdot T \cdot \mathbf{n}) \cdot \mathbf{n}$ representing normal forces (compressive or tensile), and a tangential part τ_n representing shear forces. Cutting planes exhibiting maximum shear stress (difference of the major and minor eigenvalue), are of special interest. The directions as well as the magnitude of the maximum shear are important for many failure models. Depending on the considered material, the magnitude of the tensor's deviator (Sect. 2) may be an additional indicator for the risk of failure. Together, stress and strain specify the behavior of a continuous medium under load, which allows to deduce information about a workpiece's stability.

4 Tensor Invariants

Tensor invariants are defined as scalar quantities that do not change under orthogonal coordinate transformation. In general, any scalar function of invariants will be an invariant itself. The most obvious example is the set of eigenvalues. Other common invariants are the determinant and trace or any other function that depends only on the eigenvalues.

4.1 Shape Space

The six degrees of freedom of a symmetric tensor are represented by three *direction-related* entities determining the principal directions and three eigenvalues. Following [10], we use the term *shape space* for the vector space spanned by the eigenvalues. A point in shape space is uniquely determined by the three eigenvalues. Equivalently, three coordinates according to any other reference frame (RF) of the shape-space can be used. We call these coordinates *shape descriptors*. Since a permutation of the eigenvalues describes the same tensor, it is sufficient to restrict the analysis to the *ordered subspace* where $\lambda_1 \geq \lambda_2 \geq \lambda_3$. See [1] for a closer discussion on the ordered eigenvalue space. Sometimes it is appropriate to consider further reductions of this space, e.g., if we deal with specific tensor properties, have incomplete information, or only partial interest. Depending on the property this either leads to a dimension reduction (e.g., subspace of traceless tensors) or to the definition of a subspace having the dimension of the full shape space (e.g., subset of positive-definite tensors).

An appropriate choice of RF is essential for the analysis and the understanding of a tensor in a given context. Thereby each RF provides its own view of the tensor. Characteristic tensor invariants can provide guidance for the RF specification. Considering a set of invariants as basis for the analysis of strain tensors has been proposed in the work of Criscione et al. [7]. For the analysis of diffusion tensor data, this concept has been transferred to the visualization of positive definite tensors [1, 10].

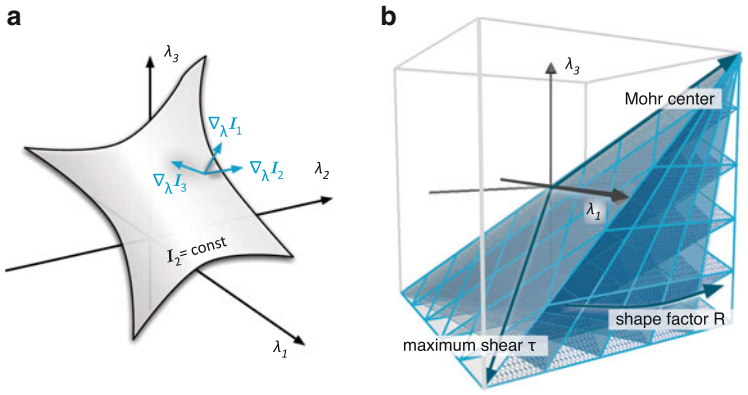


Fig. 3 (**a**) Each global invariant can be used to define 2-dimensional subspaces of the shape space. A set of three independent invariants defines a local reference frame. (**b**) Example for the parameterization of the (ordered) shape space using invariants, which are relevant for failure analysis based on the Coulomb-Mohr failure criterion

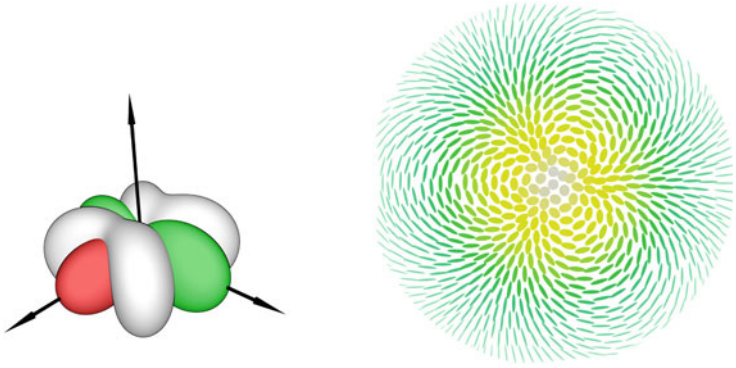
4.2 Basis Defined by Tensor Invariants

An invariant $I = f(\lambda_1, \lambda_2, \lambda_3)$ defines a family of surfaces in shape space, which are specified by a certain scalar value for I. Each set of three invariants $\{I_1, I_2, I_3\}$, which are independent, i.e., $\nabla_\lambda I_1(\nabla_\lambda I_2 \times \nabla_\lambda I_3) \neq 0$ with $\nabla_\lambda = (\partial/\partial\lambda_1, \partial/\partial\lambda_2, \partial/\partial\lambda_3)$, defines a local basis of the shape space, see Fig. 3a. A set of invariants qualifies as global basis if I_i is defined everywhere and $\nabla_\lambda I_i \neq 0$ for $i = 1, 2, 3$. From a mathematical point, of view a desirable property is orthogonality of the tensor invariants $\nabla_\lambda I_i \cdot \nabla_\lambda I_j = 0$, for all $i, j \in \{1, 2, 3\}$. In practice, however, the physical significance of tensor invariants can force us to use invariant sets that are not orthogonal. By choosing a certain basis for shape space via a set of invariants, we yield shape descriptors that provide a specific perspective onto the data.

Common reference frames for the shape space can be classified according to type, scaling, and orientation. The most important (orthogonal) types are:

- Cartesian: Full shape space.
- Spherical: Full shape space excluding the origin, where $\lambda_1 = \lambda_2 = \lambda_3 = 0$.
- Cylindrical: Full shape space excluding the symmetry axis of the cylinder.

The fundamental difference of these types lies in the nature of their coordinate types, meaning absolute (with units) respective relative (unit-free) coordinates. Relative coordinates or normalized quantities correspond to angular coordinates. They are not defined if the norm is equal to zero and, thus, are unstable for tensors with small norm. As a consequence, spherical and cylindrical coordinate systems are not optimal for indefinite tensors.



(a) Reynold + HWY Glyph (b) Uniform distribution of simple glyphs

Fig. 4 If the focus is on a specific location within the dataset, we can use complex geometries that encode as much information as possible (**a**). For a dense more continuous visualization, glyphs with simpler geometry and better perceptional properties (here ellipses) are better suited (**b**)

5 Glyphs in Tensor Visualization

Glyphs are iconic figures that encode multivariate information in terms of shape, size, color, transparency and texture. They are widely used to depict tensors. Schultz et al. [27] recently formulated the following application areas for glyphs:

- Debugging: For example, during design of new visualization methods.
- Evaluation of data quality: For example, when tensors appear as intermediate product during simulations.
- Visualization overview: For example, to get a first clue of the data and reveal patterns in it.

We would like to add *probing* to this list.

- Probing: Complex glyph geometries can be used for the detailed analysis of single tensors.

For these applications we can distinguish between dense and single-glyph visualizations (see Fig. 4). For dense glyph visualizations less complex geometries should be used. Furthermore, normalization and perceptional issues have to be considered [17, 27]. Besides the choice of an intuitive glyph the design a good placement is essential [11, 19, 21]. In the following, we concentrate on the visualization of single tensors. In order to provide easy-to-read glyphs, it is crucial to decide, *which* information to chose and *how* to map the information onto glyph parameters [23].

Which information to chose From a purely theoretical perspective, a three-dimensional tensor is perfectly represented by an ellipsoid scaled by its eigenvalues and oriented along its eigenvectors. But the ellipsoid is not capable to meet the diverse requirements imposed by the various application fields.

Many other glyph types have been presented, each with its strengths and limitations, see e.g. [13, 22, 27]. To succeed in designing glyphs that will be accepted by the domain scientist, it is important to link the question of "*which* information to use", to the choice of appropriate shape descriptors and directions. Thereby, application-specific preferences should be placed before purely theoretical considerations.

How to map the information onto geometry In general for this question there is no unique answer. Schultz et al. [27] have proposed the following general design guidelines for choosing geometries as glyphs:

- Symmetry preservation: Glyphs exhibit the symmetries as the underlying tensor.
- Continuity and disambiguity: Glyph geometries look similar for similar tensors; whereas different tensors should result in distinguishable glyphs.
- Scale invariance: Uniform scaling of the tensor by changing its norm results in a uniform scaling of the glyph geometry.
- Invariance under eigenplane projection: Projection of a tensor to a plane spanned by two eigenvectors results in a corresponding projection of the glyph.

The first two rules are very general, whereas the third rule (scale invariance) depends on the application. Especially if the tensor's deviator is of interest, glyphs that are translation invariant may be more appropriate. Therefore, we suggest to replace this rule by "invariance according to relevant shape descriptors".

6 Invariant Sets and Their Glyphs

Most common shape descriptors can be assigned to one of the coordinate-types and reference frames presented above, which represent complete, orthogonal frames. However, there are also invariant sets with high practical relevance, which are neither orthogonal nor complete. This section represents a collection of frequently used invariant sets and glyphs and their relation in different context.

6.1 Eigenvalues

Ellipsoid The most obvious invariant set are the three eigenvalues themselves. Together with the three eigenvectors they are often represented using ellipsoids. In context with stress tensors, they are also called *Lame's stress ellipsoids* or *PNS-Glyph*. In this context, the ellipsoid has an immediate physical meaning, because it displays all possible traction vectors (Eq. 1). It can be obtained by applying the tensor to the unit sphere $\{T \cdot n | n \in \mathbb{R}^3, \|n\| = 1\}$. With respect to the principal frame of reference it can also be defined as implicit surface by

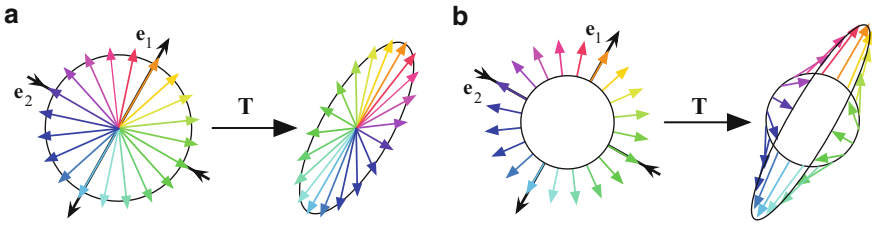
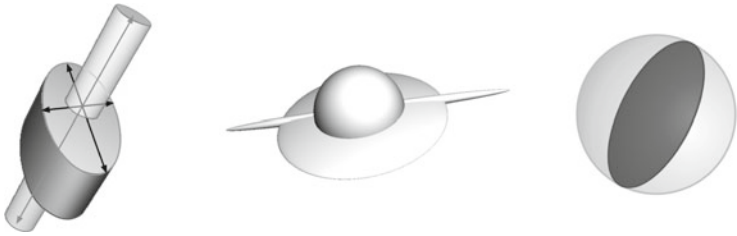


Fig. 5 Ellipse-glyphs for an indefinite 2D symmetric second-order tensor: (**a**) Applying a tensor to a unit circle results in the standard ellipse-definition, which cannot distinguish the sign of the eigenvalues. (**b**) Interpretation of the tensor as generator of a deformation of a unit sphere. The resulting glyph is the deformation ellipsoid



(a) Haber Glyph (b) Westin Glyph (c) Beachball

Fig. 6 (**a**) The Haber glyph, designed for stress tensors, emphasizes direction and magnitude of the major stress. (**b**) The major axes of the Westin glyph are aligned with the tensor's eigenvectors, emphasizing the major eigendirection. It displays the tensor invariants: linearity, planarity and isotropy. (**c**) Beachballs are a common glyph used in for earthquake visualizations. It is an example for a glyph that focuses on the representation of directions and neglects the eigenvalues

$$\frac{x_1^2}{\lambda_1^2} + \frac{x_2^2}{\lambda_2^2} + \frac{x_3^2}{\lambda_3^2} = 1. \tag{2}$$

Even though the ellipsoid represents positive-definite tensors completely, it is hard to derive any information besides the eigenvalues and eigendirections from the glyph. In addition for indefinite tensors the information about the sign of the eigenvalues is lost, see Fig. 5a. A further disadvantage of ellipsoids is visual ambiguity. Although they are of different shape, they may result in the same image after rendering, see discussion in [17].

Haber Glyph There are a variety of other glyphs focusing on the representation of the eigenvalues and principal directions. One example is the *Haber glyph* [12]. It clearly emphasizes a particular direction of interest, usually the one corresponding to the major eigenvalue. The two remaining directions are mapped to the shape of an elliptical disk with the size being controlled by the corresponding eigenvalues (see Fig. 6a).

Deformation Glyph Another glyph with the shape of an ellipsoid is the deformation ellipsoid. This glyph is capable to represent positive as well as negative

eigenvalues. It results from a deformation of the unit sphere due to the tensor, see Fig. 5b. This representation reflects the physical meaning of tensors belonging to the first class tensors, generating a change of state (Sect. 3). It can be obtained by applying the tensor to the unit sphere $\{(1+F(T))\cdot n | n \in \mathbb{R}^3, \|n\| = 1\}$. Thereby, F is an appropriate mapping technique, that is monotone in the eigenvalues and preserves their sign [15].

6.2 Principal Invariants and Trace Invariants

The *principal* or *characteristic invariants* can be found in every textbook, since they appear as coefficients of the characteristic polynomial (Sect. 2). However, they neither exhibit strong physical relevance nor nice mathematical properties. For second-order tensors, they are $I_1 = tr(\mathbf{T})$, $I_2 = 1/2(tr(\mathbf{T})^2 - tr'(\mathbf{T}^2))$, and $I_3 = \det(\mathbf{T})$. Another set of invariants, which is related and also often considered in continuum mechanics, is the *trace invariants* defined as $I = \mathrm{tr}(\mathbf{T})$, $II = \mathrm{tr}(\mathbf{T}^2)$, and $III = \mathrm{tr}(\mathbf{T}^3)$. We are not aware of any common glyph that represents these invariants.

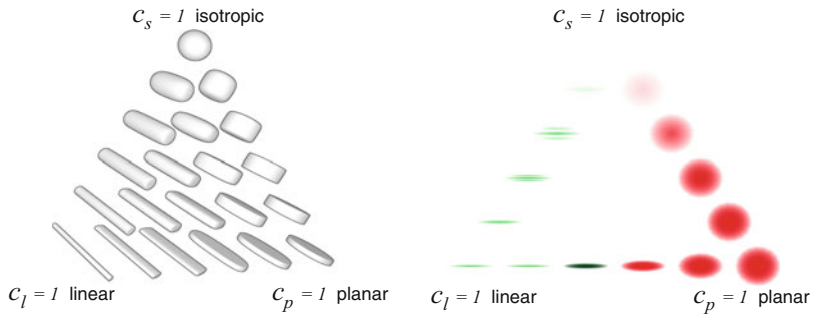
6.3 Tensor-Norm and Shape

To define scale invariant glyphs, the tensor norm is an appropriate invariant. It is often complemented by different anisotropy measures. Note, that one has to be careful when using the norm as central invariant when dealing with indefinite tensors. In general, the corresponding invariant set is not defined for the zero tensor and is often not continuously extendable to the complete shape space.

A set of complementing invariants differentiate the tensor in terms of isotropic, linear or planar anisotropic behavior [29].

$$\begin{aligned} &\text{Linear anisotropy} && c_l = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} && \in [0,1], \\ &\text{Planar anisotropy} && c_p = \frac{2(\lambda_2 - \lambda_3)}{\lambda_1 + \lambda_2 + \lambda_3} && \in [0,1], \\ &\text{Isotropy} && c_s = \frac{3\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} && \in [0,1]. \end{aligned} \tag{3}$$

These values are unit-free measures and scale invariant. They can take values from 0 to 1 and sum up to one. Thus, they are not independent and can be interpreted as barycentric coordinates, see Fig. 7. This set is especially popular for the analysis of diffusion tensors. For diffusion tensors these coordinates represent central physiological properties, which help to find regions of interest. For negative-definite



(a) Superquadrics (b) Tensor Splats

Fig. 7 Superquadrics (**a**) combine cylinders, cuboids and ellipsoids in a continuum that spans the entire range normalized tensor shapes [17]. The three extremes of linear, planar and spherical shape build the corners of a triangular barycentric space. Tensor splats (**b**) build on the same invariants equipping the ellipsoids with a texturing to enhance linearity and assigning opacity values to isotropy, assuming that isotropic tensor are of less interest (Image courtesy to Benger [3])

tensors they are not suitable, since positive and negative eigenvalues cannot be distinguished. Different glyphs have been proposed on the basis of this set of invariants. Often, they are normalized in size and, thus, do not represent the entire tensor information.

Westin glyph This glyph directly displays these shape measures using a composite shape, whose linear, planar, and spherical components are scaled accordingly [28]. Its major axes are aligned with the tensor's eigenvectors.

Superquadrics Superquadrics [2] were introduced as an extension of quadrics in order to produce a continuum of new forms by simple parameterization. Kindlmann [17] uses these superquadrics to combine base geometries in the barycentric space, spanned by the three anisotropy measures defined in Eq. 3. The base geometries for perfectly planar (a flat disk), perfectly linear (a thin cylinder) and isotropic tensors (a sphere). Recently, the *superquadric shape space* has been extended for indefinite tensors [27].

Tensor Splats Tensor splats build on the same invariants. The basic shape is an ellipsoid, which is equipped with a texture to enhance linearity. In addition opacity is assigned to isotropy, assuming that isotropic tensors are of less interest. The idea of tensor splats [3, 4] is to *splat* the glyphs onto the image plane, where they are composited in front-to-back order. For example, tensor splats can be used for DTI: Isotropic tensors are mapped to low opacity values and anisotropic tensors to high opacity values, resulting in reduced visual clutter and an emphasis on interesting regions.

Anisotropy and mode Two sets of orthogonal reference frames have been proposed and discussed in [10, 18]. The cylindrical system is an adoption of the natural strain invariants [7] into the language of DTI. It also builds on the decomposition of the tensor into its isotropic and anisotropic part. The spherical

basis uses the tensor's norm (magnitude of isotropy) as radial coordinate, mode as azimuth angle and fractional anisotropy (FA, magnitude of anisotropy) as polar angle.

$$\begin{array}{lll} \text{Frobenius norm} & R_1 = \|\mathbf{T}\| & \in [0, \infty) \\ \text{Fractional anisotropy} & R_2 = \sqrt{\frac{3}{2}} \frac{\|\mathbf{T}^{dev}\|}{\|\mathbf{T}\|} & \in [0, 1] \\ \text{Mode} & R_3 = \det(\mathbf{T}^{dev}/\|\mathbf{T}^{dev}\|) & \in [-1, 1] \end{array} \tag{4}$$

$\mathbf{T}^{dev}$ is the deviator of $\mathbf{T}$ and $\|.\|$ is the Frobenius norm. In this context, mode is used to distinguish the three basis shapes of diffusion tensors (Eq. 3): linear (mode($\mathbf{T}$) $= 1$), planar (mode($\mathbf{T}$) $= -1$) and spherical (mode($\mathbf{T}$) $= 0$). These sets have the nice property of providing an orthogonal basis of the positive definite shape-space. To our knowledge, there is no common glyph directly encoding theses measures.

Traceless tensors For symmetric traceless tensors a special variant of superquadric glyphs have been introduced in [16] to visualize the orientation of liquid crystals.

6.4 Invariant Sets for Stress Tensors

Stability and failure analysis are in the center of many questions related to stress tensors. Depending on the respective materials, considering elastic or plastic deformation, different failure models exist. For each failure model there is a specific set of invariants corresponding to the material properties. Stress-glyphs that are successfully used by domain-experts often encode exactly these sets. In the following, we give an overview over the most frequently used glyphs in this context.

Mohr's circle Many failure models build on the analysis of the maximum shear stress. It represents the maximum shear forces that occur on any cutting plane (Eq. 1). Shear forces are especially important when analyzing the failure of ductile materials. An example is the *Coulomb-Mohr failure criterion*. Its most important invariants are [24]:

$$\begin{array}{lll} \text{Maximum shear stress} & \tau = \frac{\lambda_1 - \lambda_3}{2} & \in [0, \infty) \\ \text{Mohr center} & c = \frac{\lambda_1 + \lambda_3}{2} & \in (-\infty, \infty) \\ \text{Shape factor} & R = \frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_3} & \in [0, 1] \end{array} \tag{5}$$

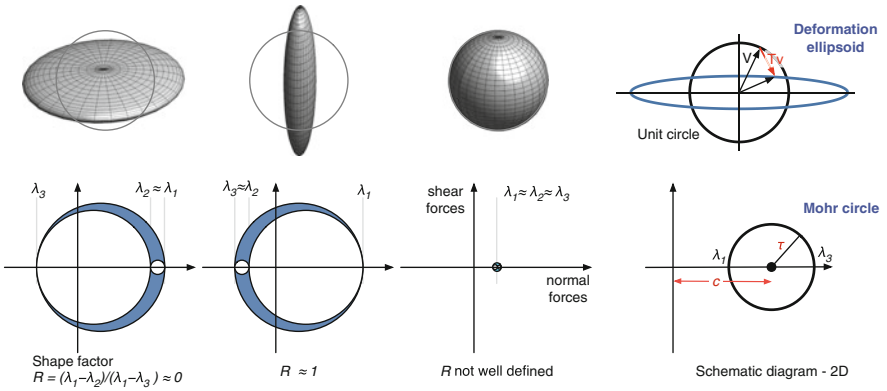


Fig. 8 Representation of the deformation ellipsoid (*first row*) of the deviator in comparison with Mohr's circle (*second row*). The horizontal axis depicts the normal stress and the vertical axis the shear stress. The *blue shaded area* represents all possible combinations of normal and shear forces for a given cutting plane. Each point within this region then corresponds to one orientation of the plane's normal. Adding a uniform pressure changes the position of Mohr's circle on the horizontal axis. Its shape remains unchanged (translation-invariance)

This set does not define an orthogonal coordinate frame but it represents important physical quantities. τ is an anisotropy measure, which has the unit of pressure. R is a relative entity and is not defined for isotropic tensors.

A typical glyph that encodes these invariants is *Mohr's circle*, which can be found in many textbooks of continuum mechanics. For visualization purposes, it has been used for stress tensors [8] as well as for diffusion tensors [5]. Mohr's circle is a two-dimensional representation of a stress tensor, which consists of three circles, see Fig. 8. It displays all possible (σ_n, τ_n)-combinations of normal σ_n and shear stresses τ_n. The radius of the largest circle displays the maximum possible shear stress. Circles that degenerate to single points represent isotropic pressure.

Mohr's circle does not encode any directional information. The glyph is translation invariant: when adding a constant multiple of the unit tensor the shape does not change. It just moves the Mohr's center along the τ_n-axis.

Note that some failure models do not consider isotropic pressure, then it is sufficient to display the Mohr's circle without specifying its center. The remaining two-dimensional shape space only considers the tensor's deviator, see Fig. 9.

Beachball In contrast, *beach balls* focus on directional information, which have an immediate physical interpretation, see Fig. 6c. They are used in geology to visualize the moment tensor describing sources of earthquakes. This application is an example where generally the entire tensor information is not available. It is often not possible to extract the part related to volume changes from the measurement data. For the definition of the beach ball it is assumed that the intermediate eigenvalue equals zero and the major and minor eigenvalue sum up to zero. The directions of interest are the tensional (positive) T- and

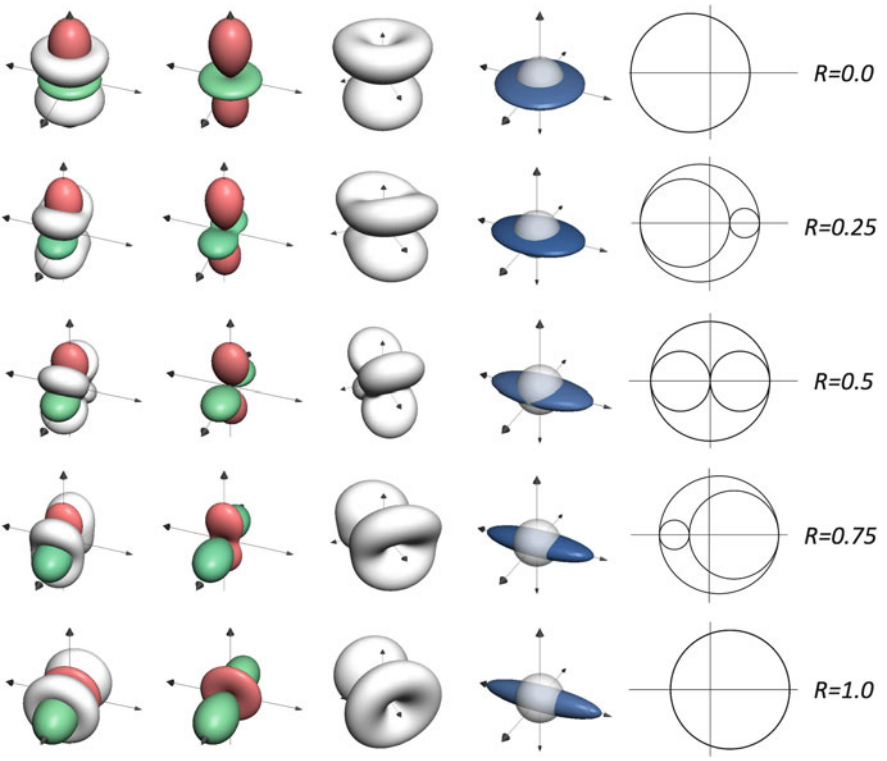


Fig. 9 Different glyphs only representing the shape factor (Eq. 5). The columns use from *left* to *right*: HWY + Reynolds, Reynolds, HWY, deformation, Mohr Glyphs. The Reynolds glyphs show directions of maximum normal forces, while the HWY glyphs highlight directions of maximum shear. For the values of $R = 0$ and $R = 1$, these directions span a cone. For all other values of R there are exactly two distinguished directions, which exhibit maximum shear

compressional (negative) P-axis, which are represented by the orientation of the beach ball.

Reynolds glyph The Reynolds glyph highlights the distribution of normal stresses σ_n depending on the normal of the cutting plane (Eq. 1). This distribution is not given by one specific invariant. The basic shape of the glyph allows to distinguish definite and indefinite tensors [22,25] (Fig. 10, second row). The glyph is defined as the set of all normal directions scaled by the magnitude of the normal stresses in that direction $\{\sigma_n \cdot \mathbf{n} | \mathbf{n} \in \mathbb{R}^3, \|\mathbf{n}\| = 1\}$. The principal directions can be depicted as directions of maximum respective minimum extension of the glyph.

HWY glyph The HWY glyph focuses on the magnitude of shear stress and can be interpreted as a counterpart to the Reynolds glyph [13]. Its displays all possible shear forces τ_n. The glyph is defined as the set of all normal directions scaled by the magnitude of the shear stresses acting in the plane perpendicular to the

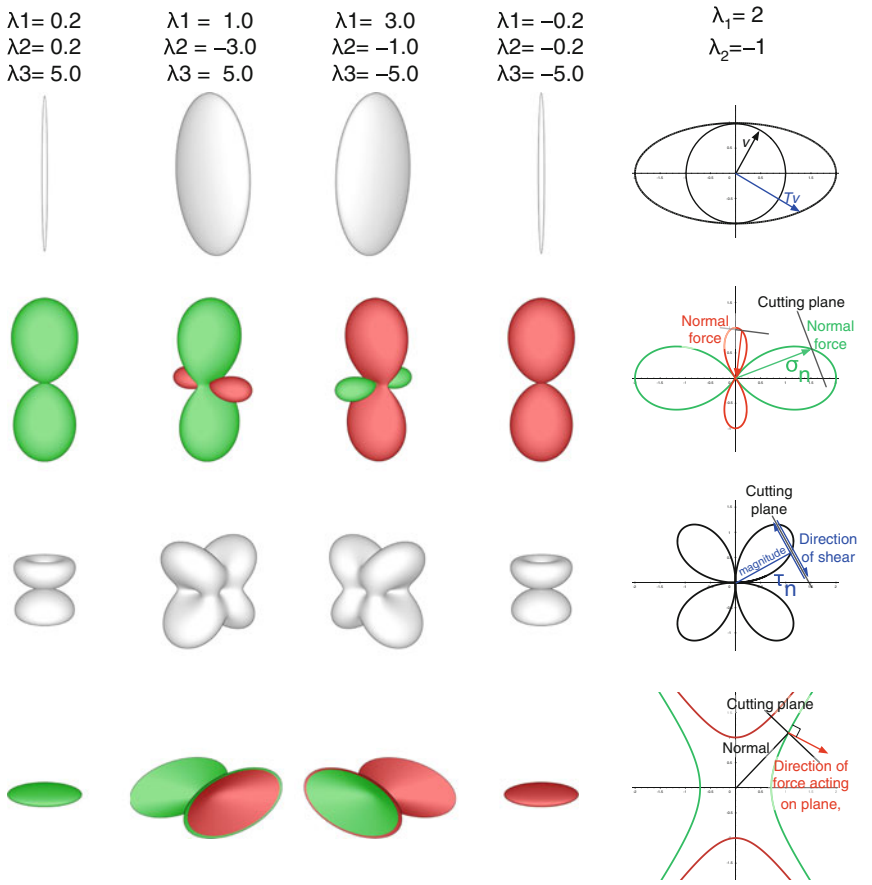


Fig. 10 Overview of basic stress glyph shapes. From top to bottom: ellipsoid (Lamé's stress ellipsoid), Reynold's glyph displaying the normal forces, HWY glyph displaying the magnitude of shear forces, and quadric surfaces. Whenever it is meaningful to distinguish compressive (*red*) and tensile (*green*) force they are colored respectively. The right column illustrates the interpretation of the various glyph types for the two-dimensional case

normal $\mathbf{n}$ $\{\tau_n \cdot \mathbf{n} | \mathbf{n} \in \mathbb{R}^3, \|\mathbf{n}\| = 1\}$ (Fig. 10, third row). The direction of the shear stress in that plane is not represented by the HWY glyph.

Quadric surface A quadric surface represents the tensor completely. It is defined as $\mathbf{T}(\underline{x}, \underline{x}) = \pm 1$, which is in terms of the eigenvalues $\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 \pm 1 = 0$, (Fig. 10, forth row). It constitutes a surface of constant distance to the center and, thus, provides a natural visualization of the metric tensor. Quadric surfaces can also be used to depict the force directions on a cutting plane and to distinguish between positive and negative eigenvalues. However, they are more complex and, hence, more difficult to comprehend. Note that the resulting surfaces are not bounded for the case of indefinite tensors, which limits their applicability.

In context with the curvature tensor of a surface there is a close correspondence to the Dupin indicatrix characterizing the local shape of a surface. It is defined as the intersection of the surface with a plane parallel to the tangent plane in a small distance away, see e.g. [9].

6.5 Invariant Set of the Natural Strain

A physically expressive cylindrical orthogonal set of invariants of the natural strain tensor was introduced by Criscione et al. [7]. The work aimed to simplify the constitutive equations, which relate stress and strain. The invariant set that is presented in their work relates to the decomposition of the tensor into its isotropic and anisotropic part. The resulting invariants then are the norm of the tensor's isotropic part as Cartesian coordinate, the norm of the deviator (measure for the degree of anisotropy) as cylindrical radius and the type of distortion, introduced as *mode*, as azimuthal angle.

$$\begin{array}{lll} \text{Magnitude/sign of dilation} & K_1 = \mathrm{tr}(\mathbf{T}) & \in (-\infty, \infty) \\ \text{Magnitude of distortion} & K_2 = \|\mathbf{T}^{dev}\| & \in [0, \infty) \\ \text{Mode} & K_3 = \dfrac{\det(\mathbf{T}^{dev})}{\|\mathbf{T}^{dev}\|} & \in [-1, 1] \end{array} \tag{6}$$

$\mathbf{D}^{dev}$ is the deviator of $\mathbf{T}$ and $\|.\|$ is the Frobenius norm. Here, mode can be interpreted as uniaxial extension for $K_3 = 1$, uniaxial contraction for $K_3 = -1$ and pure shear for $K_3 = 0$. It is not defined for $K_2 = 0$.

7 Summary

With this chapter we provide an overview of a collection of glyphs that are used to display tensors to convey domain-specific information. In many cases, these represent a distinguished set of tensor invariants. The relation between tensor invariants and glyphs can be helpful in both directions. If there are glyphs commonly used in an application, they can point to a set of invariants, which is of special interest. In a next step, these can be used as basis for other visualization methods. On the other hand, if one is given a set of invariants new glyphs could be designed around this set.

While this chapter is restricted to symmetric second-order tensors, there are also glyphs defined for higher-order tensors, see e.g. [14, 26]. A further interesting question would be the investigation of glyphs for asymmetric tensor fields.

References

1. Bahn, M.M.: Invariant and orthonormal scalar measures derived from magnetic resonance diffusion tensor imaging. J. Magn. Reson. **141**(1), 68–77 (1999)
2. Barr, A.H.: Superquadrics and angle-preserving transformations. IEEE Comput. Graph. Appl. **1**(1), 11–23 (1981)
3. Benger, W., Hege, H.-C.: Tensor splats. In: Proceedings of Visualization and Data Analysis, San Jose, pp. 151–162 (2004)
4. Bhalerao, A., Westin, C.-F.: Tensor splats: visualising tensor fields by texture mapped volume rendering. In: Proceedings of MICCAI'03, Montréal, pp. 294–901 (2003)
5. Bilgen, M., Narayana, P.A.: Mohr diagram interpretation of anisotropic diffusion indices in MRI. Magn. Reson. Imaging **21**(5), 567–572 (2003)
6. Brannon, R.M.: Kinematics: the mathematics of deformation. http://www.mech.utah.edu/~brannon/public/Deformation.pdf (2008). Accessed 24 Feb 2014
7. Criscione, J.C., Humphrey, J.D., Douglas, A.S., Hunter, W.C.: An invariant basis for natural strain which yields orthogonal stress response terms in isotropic hyperelasticity. J. Mech. Phys. Solids **48**, 2445–2465 (2000)
8. Crossno, P., Rogers, D.H., Brannon, R.M., Coblentz, D., Fredrich, J.T.: Visualization of geologic stress perturbations using Mohr diagrams. IEEE Trans. Vis. Comput. Graph. **11**(5), 508–518 (2005)
9. DoCarmo, M.P.: Differential Geometry of Curves and Surfaces. Prentice Hall, Upper Saddle River (1976)
10. Ennis, D.B., Kindlmann, G.: Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. Magn. Reson. Med. **55**(1), 136–146 (2006)
11. Feng, L., Hotz, I., Hamann, B., Joy, K.: Anisotropic noise samples. IEEE Trans. Vis. Comput. Graph. **14**(2), 342–354 (2008)
12. Haber, R.B.: Visualization techniques for engineering mechanics. Comput. Syst. Eng. **1**(1), 37–50 (1990)
13. Hashash, Y.M.A., Yao, J.I.-C., Wotring, D.C.: Glyph and hyperstreamline representation of stress and strain tensors and material constitutive response. Int. J. Numer. Anal. Methods Geomech. **27**(7), 603–626 (2003)
14. Hlawitschka, M., Scheuermann, G.: Hot-lines – tracking lines in higher order tensor fields. In: Proceedings of IEEE Visualization (Vis'05), Minneapolis, pp. 27–34 (2005)
15. Hotz, I., Feng, L., Hagen, H., Hamann, B., Jeremic, B., Joy, K.I.: Physically based methods for tensor field visualization. In: Proceedings of IEEE Visualization (Vis'04), Austin, pp. 123–130 (2004)
16. Jankun-Kelly, T., Mehta, K.: Superellipsoid-based, real symmetric traceless tensor glyphs motivated by nematic liquid crystal alignment visualizations. IEEE Trans. Vis. Comput. Graph. (Proceedings Visualization '06) **12**, 1197–1204 (2006)
17. Kindlmann, G.: Superquadric tensor glyphs. In: Proceeding of the Joint Eurographics – IEEE TCVG Symposium on Visualization, Konstanz, pp. 147–154 (2004)
18. Kindlmann, G., Ennis, D.B., Whitaker, R.T., Westin, C.-F.: Diffusion tensor analysis with invariant gradients and rotation tangents. IEEE Trans. Med. Imaging **26**, 1483–1499 (2007)
19. Kindlmann, G., Westin, C.-F.: Diffusion tensor visualization with glyph packing. IEEE Trans. Vis. Comput. Graph. (Vis '06) **12**(5), 1329–1336 (2006)
20. Kratz, A., Auer, C., Stommel, M., Hotz, I.: Visualization and analysis of second-order tensors: moving beyond the symmetric positive-definite case. Comput. Graph. Forum – State of the Art Reports **32**(1), 49–74 (2013)
21. Kratz, A., Baum, D., Hotz, I.: Anisotropic sampling of planar and two-manifold domains for texture generation and glyph distribution. Trans. Vis. Comput. Graph. (TVCG) **19**, 1782–1794 (2013)

22. Kriz, R., Yaman, M., Harting, M., Ray, A.: Visualization of zeroth, second, fourth, higher order tensors, and invariance of tensor equations. Summer Research Project 2003, 2005. University Visualization and Animation Group, Virginia Tech, Blacksburg
23. Lie, A.E., Kehrer, J., Hauser, H.: Critical design and realization aspects of glyph-based 3d data visualization. In: Proceedings of the Spring Conference on Computer Graphics (SCCG '09), Budmerice, pp. 27–34 (2009)
24. Lund, B.: Crustal stress studies using microearthquakes and boreholes. PhD thesis, Department of Earth Sciences, Uppsala University (2000)
25. Moore, J.G., Schorn, S.A., Moore, J.: Methods of classical mechanics applied to turbulence stresses in a tip leakage vortex. J. Turbonmachinery **118**(4), 622–630 (1996)
26. Schultz, T., Kindlmann, G.: A maximum enhancing higher-order tensor glyph. Comput. Graph. Forum (EuroVis'10) **29**(3), 1143–1152 (2010)
27. Schultz, T., Kindlmann, G.L.: Superquadric glyphs for symmetric second-order tensors. IEEE Trans. Vis. Comput. Graph. (Vis'10) **16**(6), 1595–1604 (2010)
28. Westin, C.-F., Maier, S.E., Khidhir, B., Everett, P., Jolesz, F.A., Kikinis, R.: Image processing for diffusion tensor magnetic resonance imaging. In: Proceedings of MICCAI'99, Cambridge, pp. 441–452 (1999)
29. Westin, C.-F., Peled, S., Gudbjartsson, H., Kikinis, R., Jolesz, F.A.: Geometrical diffusion measures for MRI from tensor basis analysis. In: Proceedings of the International Society for Magnetic Resonance Medicine (ISMRM), Vancouver, pp. 1742 (1997)

Part II
Representation and Processing of Higher-Order Descriptors

Monomial Phase: A Matrix Representation of Local Phase

Hans Knutsson and Carl-Fredrik Westin

Abstract Local phase is a powerful concept which has been successfully used in many image processing applications. For multidimensional signals the concept of phase is complex and there is no consensus on the precise meaning of phase. It is, however, accepted by all that a measure of phase implicitly carries a directional reference. We present a novel matrix representation of multidimensional phase that has a number of advantages. In contrast to previously suggested phase representations it is shown to be globally isometric for the simple signal class. The proposed phase estimation approach uses spherically separable monomial filter of orders 0, 1 and 2 which extends naturally to N dimensions. For 2-dimensional simple signals the representation has the topology of a Klein bottle. For 1-dimensional signals the new phase representation reduces to the original definition of amplitude and phase for analytic signals. Traditional phase estimation using quadrature filter pairs is based on the analytic signal concept and requires a pre-defined filter direction. The new monomial local phase representation removes this requirement by implicitly incorporating local orientation. We continue to define a phase matrix product which retains the structure of the phase matrix representation. The conjugate product gives a phase difference matrix in a manner similar to the complex conjugate product of complex numbers. Two motion estimation examples are given to demonstrate the advantages of this approach.

H. Knutsson (✉)
Department of Biomedical Engineering, Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden
e-mail: knutte@imt.liu.se

C.-F. Westin
Laboratory of Mathematics in Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Department of Biomedical Engineering, Linköping University, Linköping, Sweden
e-mail: westin@bwh.harvard.edu

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_3,

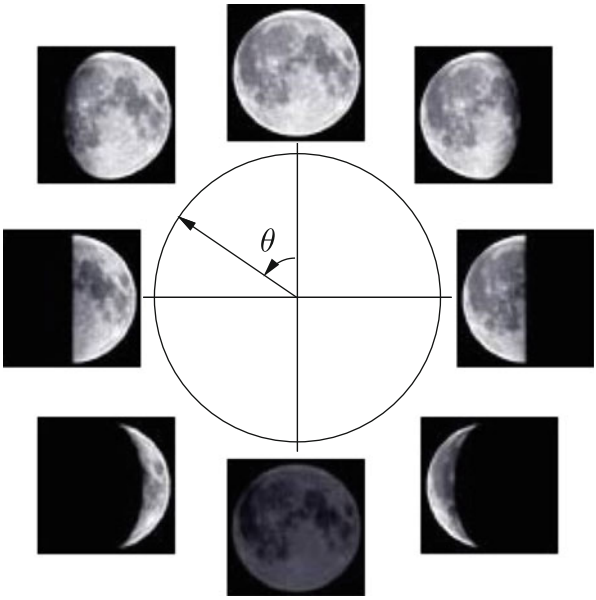


Fig. 1 The word phase is commonly used to describe the cyclic appearance of the moon. In general phase refers to a particular or possible way of viewing something, to a stage or period of change or development. In signal processing phase traditionally refers to a real number in a bounded interval, a phase angle, here denoted θ (Increasing θ show moon phases in the order they appear seen from the southern hemisphere)

1 Background

This chapter presents an overview of traditional phase concepts, and in particular discusses extension of the phase concept to encompass multi-dimensional signals.

Tracing the origin of the word "phase" one finds that it is back-formed as a singular form of modern Latin "phases", plural of the Greek word "phasis" meaning "appearance". Latin singular "phasis" was used in English from 1660. Phase is still commonly used to describe the cyclic appearance of the moon (Fig. 1). Non-lunar application is first attested 1841, meaning a difficult period in life is attested from 1913. In general phase refers to a particular or possible way of viewing something, to a stage or period of change or development.

The short introduction above is intended to underline the fact that the meaning of a concept is determined by the way it is used and that concepts will naturally evolve to best suit the communication needs of a given group of people. This is no doubt true also in the evolution of science. We will review, investigate, and further develop the concept of phase in the context of signal processing.

In signal processing the precise meaning of a concept is defined by it's mathematical representation. A meaningful application of most signal processing concepts requires that the data to be processed represent some aspect of the real world in an orderly way. More precisely, it is generally required that an increased difference between real world events results in an increased distance between the data points that represent these events. Working with representations where these requirements are not met would make many signal processing concepts meaningless and would also greatly reduce the possibility of gaining an intuitive understanding

of how suitable processing can be carried out. For many real world aspects, however, establishing well-behaved representations is a non-trivial task and in these cases a first and crucial step of any analysis is to find such a representation. In this chapter the required features of a suitable mathematical representation of local phase for N-dimensional signals are discussed and a novel matrix representation is proposed.

1.1 Traditional 1-Dimensional Signal Processing Concepts

Description and estimation of local spatio-temporal structure has a long history and numerous analysis tools have been developed. In addition to local phase, local orientation, frequency, scale, motion, and locality of estimates are prominent examples of features that have been considered central in the analysis. In the field of signal processing phase was originally used as a descriptor for one-dimensional signals. The concept of phase was later broadened to serve as a descriptor of multi-dimensional structure.

Many of the popular image analysis tools related too local phase have roots that can be traced to early work in signal processing and optics, e.g. Riesz transforms [1], Zernike moments [2, 3], and Gabor signals [4]. Regardless of this development the first stages in the analysis remain the same. In most cases the processing starts by performing local linear combinations of image values, e.g. convolution operators.

Important signal descriptors were often first developed for temporal, 1-dimensional signals. Important well-known concepts are the Fourier phase and the shift theorem, describing how the Fourier phase is affected by moving the signal. However, since the sine wave basis functions in the Fourier transform are inherently global, the Fourier phase concept is of limited practical utility. Real world signals are often non-stationary (images, volumes, sequences) and thus local features in time and space are often of more interest. The local phase concept is readily defined via local basis functions and the Hilbert transform [5–7]. The first mention of phase in a 2-dimensional image processing context appears to be found in [8]. Early work on phase in more than one dimension is found in [9–13]. Work on extensions of the Hilbert transform to higher dimensions can be found in [14]. Related work can also be found in references [15–26]. An overview of phase representation in signal processing is given in [27]. More recent relevant work on phase related topics is presented in [28–31]. Specific relations of the cited publications to the present work will be commented on in the proper context below.

1.1.1 The Hilbert Transform

At the heart of the 1-dimensional concepts of local phase, local frequency, and local amplitude lies the *Hilbert transform*. Denoting the spatial variable by x the Hilbert transform of a signal $\mathtt{s}(x)$ is obtained through the following convolution operation. Following the definition and notation of Bracewell [7], we write this as:

$$\mathsf{s}_{\mathscr{H}}(x) = \mathsf{s}(x) * \frac{-1}{\pi x} \tag{1}$$

where $*$ denotes the convolution operator.[1] Let ω denote the frequency variable and $\mathsf{S}_{\mathscr{H}}(\omega)$ denote the Fourier transform corresponding to $\mathsf{s}_{\mathscr{H}}(x)$. Then, since convolution in the spatial domain corresponds to multiplication in the Fourier domain, the Fourier transform of $\frac{-1}{\pi x}$ is $i\ \mathrm{sign}(\omega)$, can be compactly expressed in the Fourier domain:

$$\mathsf{S}_{\mathscr{H}}(\omega) = i\ \mathsf{S}(\omega)\,\mathrm{sign}(\omega) \tag{2}$$

where $\mathrm{sign}(\omega)$ equals one for $\omega > 0$, zero for $\omega = 0$ and minus one for $\omega < 0$. Hence, the Fourier transform of $\mathsf{s}_{\mathscr{H}}$ is obtained by multiplying S by the imaginary unit i and then change sign of the result for negative frequencies. Another way of understanding this transform is that the argument of the frequency components are turned an angle $\frac{\pi}{2}$ in the positive direction for positive frequencies and in the negative direction for negative frequencies.

1.1.2 The Analytic Signal

Having defined the Hilbert transform we can now define the *analytic signal*, s_A, corresponding to a real signal s. The analytic signal is a complex signal and is uniquely defined as:

$$\mathsf{s}_A(x) = \mathsf{s}(x) - i\ \mathsf{s}_{\mathscr{H}}(x) \tag{3}$$

As can be seen, the real part of the analytic signal is the original signal itself and the imaginary part is it's Hilbert transform times $-i$. This is illustrated in Figs. 2 and 5. To summarize, the relations between a real signal and it's corresponding analytic signal is, in the spatial and frequency domains given by:

$$\begin{cases} \mathsf{s}_A(x) &= [\,\delta(x) + \frac{i}{\pi x}\,] * \mathsf{s}(x) \\ \mathsf{S}_A(\omega) &= [\,1 + \mathrm{sign}(\omega)\,]\ \mathsf{S}(\omega) = 2\,H(\omega)\,F(\omega) \end{cases} \tag{4}$$

where $H(\cdot)$ is the Heavyside step function: $H(\omega) = 1$ for $u > 0$ and $H(\omega) = 0$ for $\omega < 0$. Hence, the analytic signal corresponding to s is obtained by suppressing all it's negative frequencies and multiplying by two. Two simple examples are:

[1] Some researchers may prefer to express Eq. (1) as: $f_{\mathscr{H}}(x) = (f * g)(x)$; $g(x) = \frac{-1}{\pi x}$. We will, however, continue to use the notation of Bracewell, i.e. $f_{\mathscr{H}}(x) = f(x) * g(x)$, no confusion should arise from this.

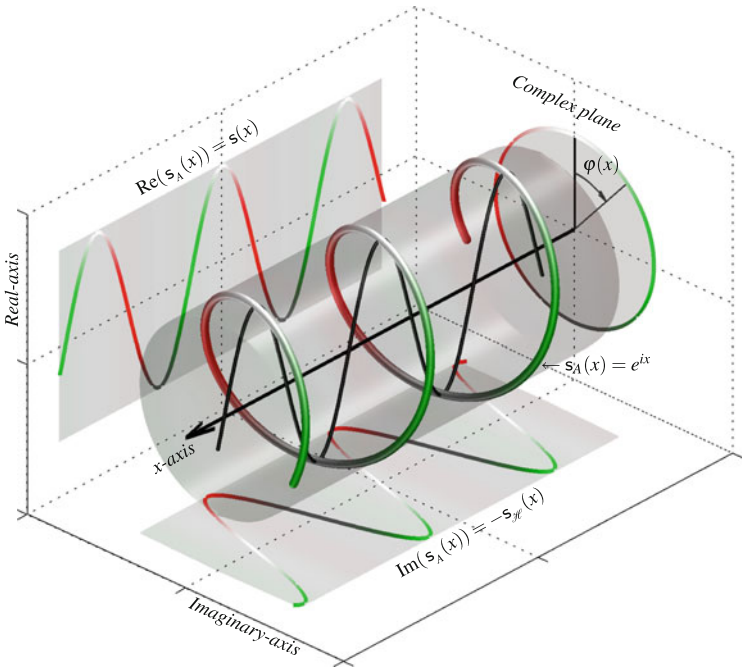


Fig. 2 Figure showing perhaps the simplest possible analytic signal. The real part is a cosine, $\mathsf{s}(x) = \cos(x)$, and the imaginary part is a sine, $\mathsf{s}_{\mathscr{H}}(x) = -\sin(x)$. The corresponding analytic signal is a complex exponential, $\mathsf{s}_A(x) = e^{ix}$. Color code: *white* means $\varphi = 0$ (zero phase), *dark gray* means $\varphi = \pi$, *green* means $\varphi = \pi/2$ and *red* means $\varphi = -\pi/2$. The *gray* 'glass' tube represents the signal amplitude. The original signal is shown in *black* in the center

$$\begin{aligned} \mathsf{s}(x) = \cos(\omega x) \quad &\leftrightarrow \quad \mathsf{s}_A(x) = \quad e^{i\omega x} \\ \mathsf{s}(x) = \sin(\omega x) \quad &\leftrightarrow \quad \mathsf{s}_A(x) = -ie^{i\omega x} \end{aligned} \tag{5}$$

These two sinusoidal examples have infinite support, the analytic signal has a constant magnitude and the difference can be describes as a shift in phase angle, see Fig. 2.

1.1.3 Analytic Signal Examples

An interesting family of analytic signals, $\mathsf{s}_A(x)$, and the corresponding Fourier transforms, $\mathsf{S}_A(\omega)$, is given by the Poisson distribution related function:

$$\begin{cases} \mathsf{s}_A(x) = (1+ix)^{-\kappa} \\ \mathsf{S}_A(\omega) = \frac{H(\omega)}{\Gamma(\kappa)}\, \omega^{\kappa-1}\, e^{-\omega} \end{cases} \tag{6}$$

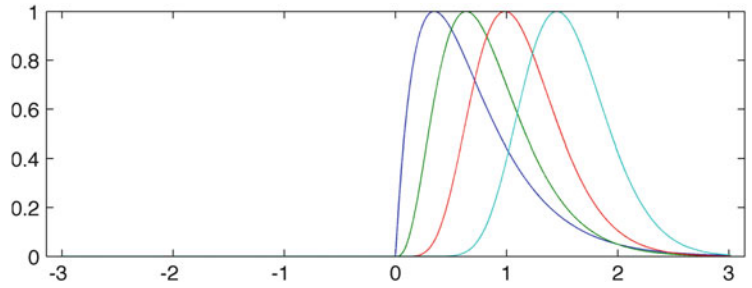


Fig. 3 Fourier domain plots of four different analytic signals (or quadrature filters) from the family of analytic functions given by Eq. (6). Regardless of the value of κ there are no negative frequencies present. From left to right the values of κ are: 2, 4, 8 and 16. The plots have been normalized to have a maximum value of one (Different colors simply indicate different curves)

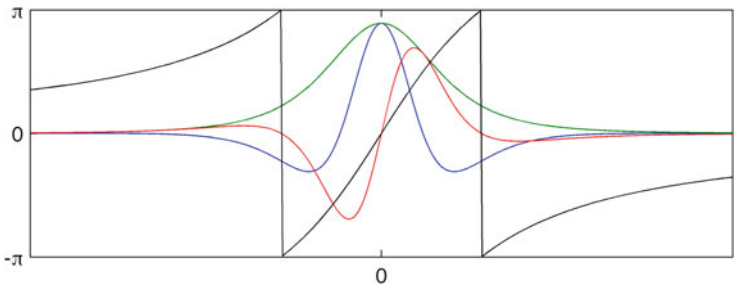


Fig. 4 Traditional plot of an analytic signal from the family given in Eq. (6) with $\kappa = 4$. The figure shows the real part (*blue*), the imaginary part (*red*), the amplitude (*green*) and the phase angle with wrap-around at $\pm\pi$ (*black*)

where $\Gamma(\cdot)$ is the gamma function and κ (> 0) is a real number determining the shape of the analytic signal. Figure 3 shows four examples from this family of functions in the Fourier domain. Figures 4 and 5 show the corresponding signal in the spatial (or temporal) domain for $\kappa = 4$. Figure 4 shows a traditional plot where the 'wrap around' discontinuity of the phase angle at $\pm\pi$ can be seen. Figure 4 shows the analytic signal in a three-dimensional space where the phase can be represented as a continuously varying vector in the complex plane, $e^{\varphi(x)}$.

1.1.4 Local Amplitude, Phase and Frequency

Three traditionally important and fundamental concepts in 1-dimensional signal processing are:

$$\begin{aligned} &\text{The local amplitude :} && \mathsf{A}(x) \\ &\text{The local phase angle :} && \varphi(x) \\ &\text{The local frequency :} && \frac{\partial \varphi(x)}{\partial x} \end{aligned} \tag{7}$$

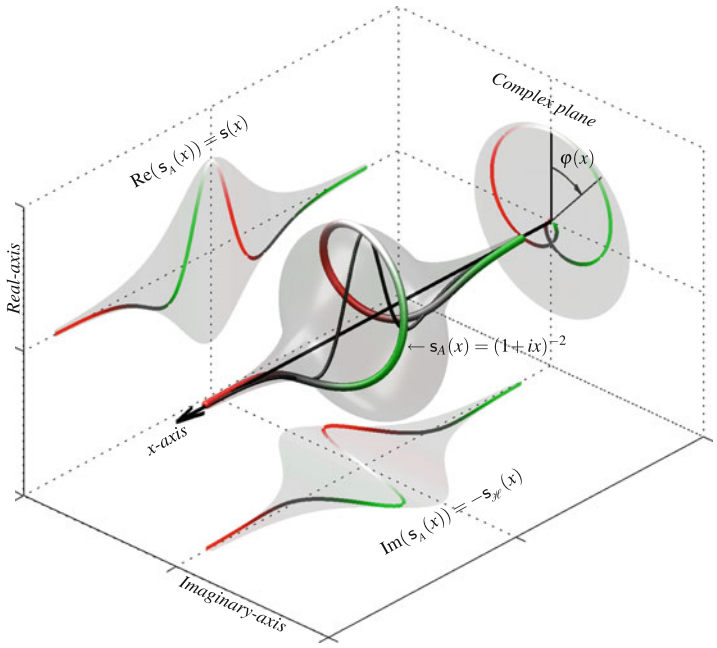


Fig. 5 Figure showing a relatively broad-band analytic signal from the family given in Eq. (6). Here $\kappa = 4$ implying that the phase angle, φ, will make two full 2π turns from start to end of the signal. Color code: *white* means $\varphi = 0$ (zero phase), *dark gray* means $\varphi = \pi$, *green* means $\varphi = \pi/2$ and *red* means $\varphi = -\pi/2$. The *gray* 'glass bulb' represents the signal amplitude. The original signal is shown in *black* in the center

An analytic signal can be directly expressed in terms of amplitude and phase:

$$s_A(x) = A(x)\, e^{i\,\varphi(x)} \tag{8}$$

In general, and in particular for analytic signals constructed from a real signal according to Eq. (3), the amplitude varies more slowly than the phase and the two concepts provide a useful complementary local representation of a signal. This way of expressing the analytic signal has found numerous applications in signal processing, [7].

1.1.5 Quadrature Filters

A filter that has the properties of an analytic signal is known as a quadrature filter. The real even part will pick up the even part of the signal and the imaginary odd part will pick up the odd part of the signal. If the desired Fourier response is only known for the even or the odd part the missing part can be generated directly in the Fourier domain by using the sign function.

$$\begin{cases} F_o(\omega) = \text{sign}(\omega)\, F_e(\omega) \\ F_e(\omega) = \text{sign}(\omega)\, F_o(\omega) \end{cases} \tag{9}$$

In the spatial domain, the even filter is real and the odd filter is imaginary and it is natural to combine them into a single complex filter, $f(x)$:

$$\begin{aligned} f(x) &= f_e(x) + f_o(x) \\ &= m(x)\, e^{i\,\theta(x)} \end{aligned} \tag{10}$$

Equation (10) shows that the filter can be expressed as a product of the filter magnitude, m, and a unitary complex number, $e^{i\,\theta(x)}$. The argument (modulo 2π) of the latter is traditionally referred to as the phase. It is, however, from a representational point of view much preferable to use the unitary complex number, $\psi = e^{i\,\theta(x)}$, as the representation of phase since it is continuous and does not suffer from the wrap-around discontinuity. This will also be consistent with the representation for higher dimensional phase suggested below. When referring to the real argument, θ, we will use the term 'phase angle'.

Since the Hilbert transform is defined in one dimension the analytic signal is only well-defined for one-dimensional signals as well. The Hilbert transform can, however, be used in higher dimensional spaces if a direction in this space is specified. Local phase estimation based on such directed quadrature filter responses [8, 17] has found extensive use in image processing. The quadrature filter response to a signal, s, can be expressed as a convolution in the spatial domain or as a multiplication in the Fourier domain, i.e.

$$\begin{aligned} \mathsf{q}(x) &= f(x) * \mathsf{s}(x) \\ &= \mathscr{F}^{-1}[\, F(\omega)\, \mathsf{S}(\omega)\,] \end{aligned} \tag{11}$$

where $*$ denotes the convolution operator, S denotes the Fourier transform of s, $\mathscr{F}^{-1}$ denotes the inverse Fourier transform. The quadrature filter response can also be represented in terms of amplitude, A and phase angle θ or phase ψ:

$$\begin{aligned} \mathsf{q}(x) &= \mathsf{A}(x)\, e^{i\,\theta(x)} \\ &= \mathsf{A}(x)\psi(x) \end{aligned} \tag{12}$$

The magnitude of the filter response reflects the signal energy and the argument reflects the relationship between the evenness and oddness of the signal (see Fig. 6).

1.1.6 Interpretation of Local Phase

The fundamental property of local phase represents the relation between even and odd signal content at specific spatial location. The local phase has a number of

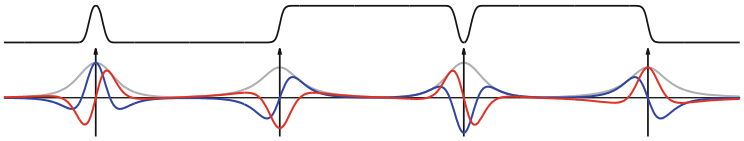


Fig. 6 Plots of a quadrature filter response to a signal (*top*) containing four typical events: a positive 'bump', a positive gradient edge, a a negative 'bump' and a negative gradient edge. The real response is shown in *blue*, the imaginary in *red* and the amplitude in *gray*

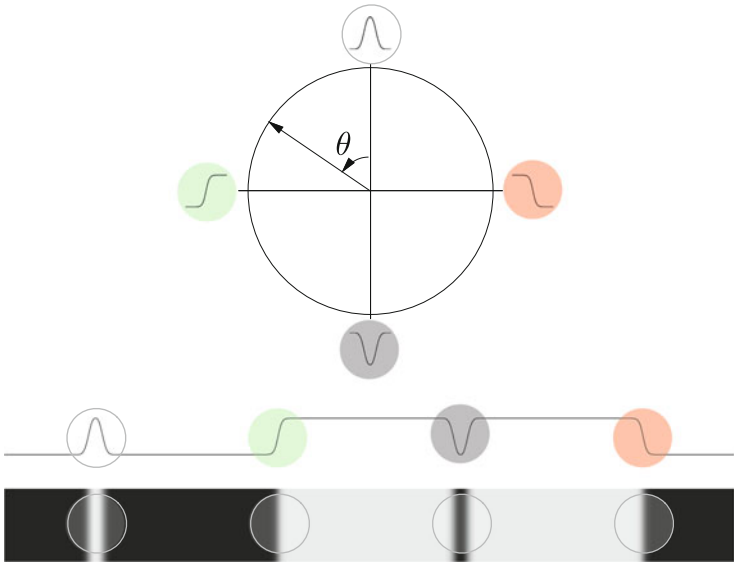


Fig. 7 A simple display of the 1-dimensional phase concept. *Bottom*: an image with varying intensity in the x-direction. Marked with *circles* are from left to right: a *white line*, a positive gradient edge, a *black line* and a negative gradient edge. *Middle*: the corresponding intensity as a 1-dimensional signal. *Top*: the local signal shapes placed at the corresponding phase angle positions

interesting invariance and equivariance properties that makes it an important feature in image processing. Local phase estimates are, for example, invariant to signal energy, the phase varies in the same manner regardless if there are small or large signal variations. Further, local phase and spatial position have a tight relationship and local phase generally varies continuously with spatial position thus enabling sub-pixel/voxel resolution. In high frequency areas the phase changes faster than in low frequency areas. The phase angle derivative is called local or instantaneous frequency [7, 16]. Phase also has interesting invariance properties with respect to scaling [9–11]. Figure 7 shows the intensity profile over a number of lines and edges.

In this chapter we will show how the concept of phase can be generalized to higher dimensional signal spaces.

2 Phase Representation for Multi-dimensional Signals

For multidimensional signals the equivalent of the 1-dimensional quadrature filter response amplitude is the structure tensor [15, 29]. Structure tensors are, just as the response amplitude, phase invariant. However, the relationships between the different filters used to produce the structure tensor holds important complementary information about the neighborhood. These relations are the basis for the concept known as local phase. Local phase is a powerful concept in it's own right and has been successfully used in many applications. As is the case with the structure tensor there is no consensus on the precise meaning of phase. In this chapter we will make an effort to define what the concept of phase should imply.

2.1 General Representational Considerations

At a high level of abstraction a representation of any concept can be viewed in terms of equivariance and invariance. A good representation should vary in a way that precisely reflects changes in the feature that is represented. At the same time the representation should be invariant to changes in features that are considered to leave that which is represented unchanged. The common understanding is that local phase should be equivariant to a relation between oddness and evenness of the signal at a given position. It is also accepted by all that a measure of phase necessarily carries a directional reference, i.e. it should be equivariant with rotation of the signal. That the phase of a signal should be invariant to the signal mean level and the signal amplitude is also not debated. We will, in addition, consider the following properties to apply to a good phase representation for multi-dimensional signals:

- Uniqueness
- Continuity
- A shift invariant manifold metric

The purpose of these requirements is to ensure that common signal processing operations, such as averaging and differentiation, can be meaningfully applied to a spatio-temporal phase field. Precisely what the requirements are taken to imply will be made clear in a proper mathematical setting in Sect. 3.

2.2 Monomial Filters

All local phase estimation approaches are based on the use of a set of filters onto which each local neighborhood is projected. The design of these filters directly determines crucial aspects of the performance of the estimator. The filters should provide an appropriate basis for representing the targeted features of the signal.

A natural requirement is that the estimate directly reflects rotations of the neighborhood, i.e. the estimate should be equivariant with rotation, but be invariant to other transformations, e.g. change of scale. Another important aspect, not further discussed here, is the locality of the estimates, see [23].

Here we present a class of filters designed to meet the above requirements, *monomial filters*. The monomial filters are spherically separable, i.e. defined as a product of one radial and one directional part:

$$F(\boldsymbol{\omega}) = R(\rho)\, D(\hat{\boldsymbol{\omega}}) \tag{13}$$

where $\boldsymbol{\omega}$ is the Fourier domain coordinate, $\rho = \|\boldsymbol{\omega}\|$ and $\hat{\boldsymbol{\omega}} = \frac{\boldsymbol{\omega}}{\rho}$.

2.2.1 Radial Part

The radial part is not important for the following discussions on phase and could in principle be set to unity. In practice, however, there will always be a non-constant radial part involved, typically it is a bandpass filter ($R(0) = 0$), e.g. a *lognormal* function, [8, 17], a *logerf* function, [23], or a function given by $|\mathsf{S}_A(\omega)|$ in Eq. (6).

2.2.2 Directional Matrix

The directional part consists of monomials i.e. products of non-negative integer powers of the components of $\hat{\boldsymbol{\omega}}$. Performing n repeated outer products of $\hat{\boldsymbol{\omega}}$ will contain all order n component products.

$$\boldsymbol{D}_n(\hat{\boldsymbol{\omega}}) = \hat{\boldsymbol{\omega}}^{\otimes n} = \underbrace{\hat{\boldsymbol{\omega}} \otimes \hat{\boldsymbol{\omega}} \ \ldots \otimes \hat{\boldsymbol{\omega}}}_{n \text{ entities}} \tag{14}$$

For the following investigation of phase only orders 0, 1 and 2 will be needed. In the 2D case, using the notation $\hat{\boldsymbol{\omega}} = (\mu,\ \nu)^T$, we have:

$$D_0(\hat{\boldsymbol{\omega}}) = 1 \tag{15}$$

$$\boldsymbol{D}_1(\hat{\boldsymbol{\omega}}) = \begin{pmatrix} \mu \\ \nu \end{pmatrix} \tag{16}$$

$$\boldsymbol{D}_2(\hat{\boldsymbol{\omega}}) = \begin{pmatrix} \mu^2 & \mu\nu \\ \mu\nu & \nu^2 \end{pmatrix} \tag{17}$$

It is worth noting here that $\boldsymbol{D}_1(\hat{\boldsymbol{\omega}})$ corresponds to the Riesz transform, [1] in the general case and the Hilbert transform in the one-dimensional case since $\hat{\boldsymbol{\omega}} = \frac{\omega}{|\omega|} = \text{sign}(\boldsymbol{\omega})$.

2.2.3 Monomial Filter Matrices

To construct matrices holding proper filters we simply multiply the matrix holding the directional responses, $\boldsymbol{D}_n(\hat{\boldsymbol{\omega}})$, with the radial function, $R(\rho)$. For each order $n \geq 0$ a *monomial filter matrix* is defined as:

$$\boldsymbol{F}_n(\boldsymbol{\omega}) = R(\rho)\ \ \boldsymbol{D}_n(\hat{\boldsymbol{\omega}}) \tag{18}$$

2.3 Monomial Filter Response Matrices

The next step is to apply a monomial filter matrix to a signal, thus obtaining a *monomial filter response matrix*. This can be done by convolving the signal with each of the filters in the filter matrix and storing the results in the corresponding positions. Using an FFT approach the same result can also be obtained multiplying the Fourier transform of the signal by each filter in the Fourier domain.

Let the spatial domain correspondence of the monomial filter matrix $\boldsymbol{F}_n$ be denoted $\mathbf{F}_n$. Each element of $\mathbf{F}_n$ contains the convolution kernel of the corresponding FD filter function in $\boldsymbol{F}_n$. If the multi-dimensional signal is denoted $\mathsf{s}(\mathbf{x})$, where $\mathbf{x}$ denotes the spatial coordinates, the monomial filter response matrix, $\mathbf{Q}_n(\mathbf{x})$, is defined as:

$$\mathbf{Q}_n(\mathbf{x}) = \mathbf{F}_n(\mathbf{x}) * \mathsf{s}(\mathbf{x}) \tag{19}$$

where $*$ denotes the convolution operator.

Denoting the Fourier transform of s by S the same relation can be expressed using multiplication in the Fourier domain:

$$\mathbf{Q}_n(\mathbf{x}) = \mathscr{F}^{-1}[\boldsymbol{F}_n(\boldsymbol{\omega})\ \ \mathsf{S}(\boldsymbol{\omega})\,] \tag{20}$$

Here $\mathscr{F}^{-1}$ denotes the inverse Fourier transform.

In this general description each element of $\mathbf{Q}_n(\mathbf{x})$ contains the monomial filter responses for the entire signal. Since all filtering operations in this paper are shift invariant we may, in the interest of clarity and without loss of generality, from now on omit to denote the spatial coordinate vector $\mathbf{x}$ and only consider the filter matrix response at $\mathbf{x} = \mathbf{0}$.

2.4 Signal Classes

It will be useful for the following discussion to define two different classes of signals. We will here define the sinusoidal and simple signal classes.

2.4.1 Sinusoidal Signals

We first present the simplest possible case, a sinusoidal signal with amplitude A, spatial frequency $\mathbf{u}$, and zero phase.

$$\mathsf{s}(\mathbf{x}) = \mathsf{A}\cos(\mathbf{u}^{\mathrm{T}}\mathbf{x}) \tag{21}$$

For this case the monomial filter response matrix dependence on the signal frequency, $\mathbf{u}$, is given by:

$$\mathbf{Q}_n(\mathbf{u}) = \mathsf{A}\, R(\rho)\, \boldsymbol{D}_n(\hat{\mathbf{u}}) \qquad \text{for even } n \tag{22}$$

where $\rho = \|\mathbf{u}\|$ and $\hat{\mathbf{u}} = \frac{\mathbf{u}}{\rho}$. Note that in this special case of zero phase, i.e. a symmetric signal, the response will be zero for odd n. For odd signals, i.e.

$$\mathsf{s}(\mathbf{x}) = \mathsf{A}\sin(\mathbf{u}^{\mathrm{T}}\mathbf{x}) \tag{23}$$

even orders will be 0 and odd order responses are given by:

$$\mathbf{Q}_n(\mathbf{u}) = -i\,\mathsf{A}\, R(\rho)\, \boldsymbol{D}_n(\hat{\mathbf{u}}) \qquad \text{for odd } n \tag{24}$$

For a general sinusoidal with phase θ, i.e.

$$\mathsf{s}(\mathbf{x}) = \mathsf{A}\cos(\mathbf{u}^{\mathrm{T}}\mathbf{x} + \theta) \tag{25}$$

both even and odd order filters will respond and we get:

$$\mathbf{Q}_n(\theta, \mathbf{u}), = \begin{cases} \mathsf{A}\cos(\theta)\, R(\rho)\, \boldsymbol{D}_n(\hat{\mathbf{u}}) & \text{for even } n \\ -i\,\mathsf{A}\sin(\theta)\, R(\rho)\, \boldsymbol{D}_n(\hat{\mathbf{u}}) & \text{for odd } n \end{cases} \tag{26}$$

2.4.2 Simple Signals

Following [17] we define simple signals to be signals that can be expressed as:

$$\mathsf{s}(\mathbf{x}) = \mathsf{g}(\hat{\mathbf{u}}^{\mathrm{T}}\mathbf{x}) \tag{27}$$

Here $\mathsf{g}(\cdot)$ is any real one variable function and $\hat{\mathbf{u}}$ is a unit vector giving the orientation of the signal. For this case the monomial filter response matrix dependence on the signal orientation, $\hat{\mathbf{u}}$, is given by:

$$\mathbf{Q}_n = \mathsf{A}_n\, \boldsymbol{D}_n(\hat{\mathbf{u}}) \tag{28}$$

Here A_n is the local amplitude of the filter response. By factoring out the directional dependence, A_n depends only on the radial filter function, $R(\rho)$, and the signal

generating function, $\mathsf{g}(\mathsf{x})$. By he Fourier slice theorem, [7], we know that the Fourier transform of a simple signal is non-zero only on a line through the origin. This makes for a simple solution. Letting v be a 1-dimensional frequency variable and denoting the Fourier transform of $\mathsf{g}(x)$ by $\mathsf{G}(\mathsf{v})$ we find the filter response amplitude as:

$$\begin{cases} \mathsf{A}_n = \mathsf{A}_\mathsf{e} = \int_v R(|v|)\, \mathsf{G}(v)\, dv & \text{for even } n \\ \mathsf{A}_n = \mathsf{A}_\mathsf{o} = \int_v R(|v|)\, \mathsf{G}(v)\, \text{sign}(v)\, dv & \text{for odd } n \end{cases} \tag{29}$$

Unless explicitly mentioned all signals will in the following be regarded simple.

2.5 Vector Phase Representations

Next we turn to a somewhat more advanced phase representation approach. Monomial quadrature filter sets support a simple yet general definition of phase. We will show how a continuous and consistent $N+1$ dimensional vector phase representation can be constructed.

2.5.1 Monomial Quadrature Filter Matrices

The final step required for attaining filter matrices that can be used for local phase estimation is to concatenate one even and one odd monomial filter matrix to form a monomial quadrature filter matrix, $\boldsymbol{F}_{mn}(\boldsymbol{\omega})$.

$$\boldsymbol{F}_{mn}(\boldsymbol{\omega}) = \{\boldsymbol{F}_m(\boldsymbol{\omega}),\ \boldsymbol{F}_\mathsf{n}(\boldsymbol{\omega})\} \tag{30}$$

Where the "{}" brackets denotes a simple juxtaposition of the components, i.e. there are no requirements on the types or sizes of the assembled components. The monomial quadrature filter response matrix, $\mathbf{Q}_{nm}$, is, as before, obtained through convolution in the spatial domain or multiplication in the Fourier domain and we have:

$$\mathbf{Q}_{mn} = \{\mathbf{Q}_m,\ \mathbf{Q}_n\} \tag{31}$$

2.5.2 Phase from Orders 0 and 1

The simplest case of phase estimation uses filters of orders 0 and 1, i.e.

$$\begin{aligned} \boldsymbol{F}_{01}(\mathbf{u}) &= \{\boldsymbol{F}_0(\mathbf{u})\ ,\ \boldsymbol{F}_1(\mathbf{u})\} \\ &= R(\rho)\,\{1\ ,\ \hat{\mathbf{u}}\} \end{aligned} \tag{32}$$

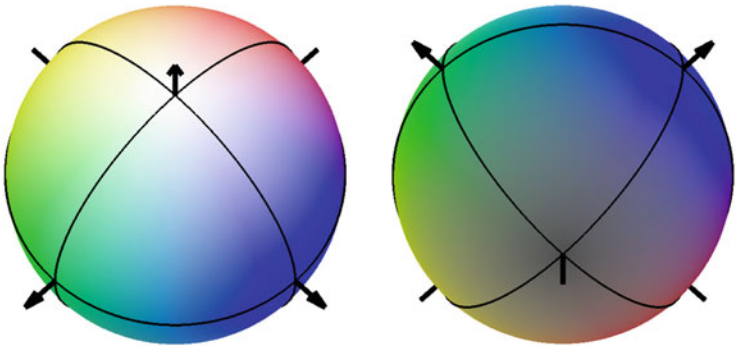


Fig. 8 Two views (from above and from below) of the color coded 3-dimensional vector phase space for 2-dimensional signals. The color code is as follows: *white* means $\theta = 0$, increased saturation means higher θ and full saturation means $\theta = \pi/2$. *Gray* means $\theta = \pi$. The color indicates the direction, φ. *Green/red* indicates x-direction. *Blue/yellow* indicates y-direction. *Black* iso-line are drawn for $\theta = pi/4, \pi/2, 3\pi/2$ and for $\varphi = 0, pi/2$. *Right*: the same color coded space seen from below

The filter response using $\mathbf{F}_{01}$, the spatial counterpart of $\boldsymbol{F}_{01}$, can then be written:

$$\begin{aligned} \mathbf{Q}_{01} &= \mathbf{F}_{01} * \mathsf{s} \\ &= \{\, \mathsf{A}_e \;,\; \mathsf{A}_o \; \hat{\mathbf{u}} \,\} \end{aligned} \tag{33}$$

A phase representation, $\boldsymbol{\theta}_{01}$, is then obtained by simply normalizing the filter response matrix, $\mathbf{Q}_{01}$, using the Frobenius norm.

$$\begin{aligned} \boldsymbol{\Psi}_{01} \equiv \hat{\mathbf{Q}}_{01} &= \frac{\mathbf{Q}_{01}}{\|\mathbf{Q}_{01}\|} \\ &= \frac{\{\, \mathsf{A}_e \;,\; \mathsf{A}_o \; \hat{\mathbf{u}} \,\}}{\sqrt{\mathsf{A}_e^2 + \mathsf{A}_o^2}} \\ &= \{\cos(\theta) \;,\; i \; \sin(\theta) \; \hat{\mathbf{u}} \,\} \end{aligned} \tag{34}$$

A color coded representation of the N + 1 dimensional phase for 2-dimensional signals is shown in Fig. 8. The eight plots in Fig. 9 show the local phase when moving in two different directions in a gray scale image using three different types of visualizations.

2.6 Relations to Previous Work

Before moving on to discussing higher order phase representations we would like to point out a number of interesting relations to classical concepts and previous work.

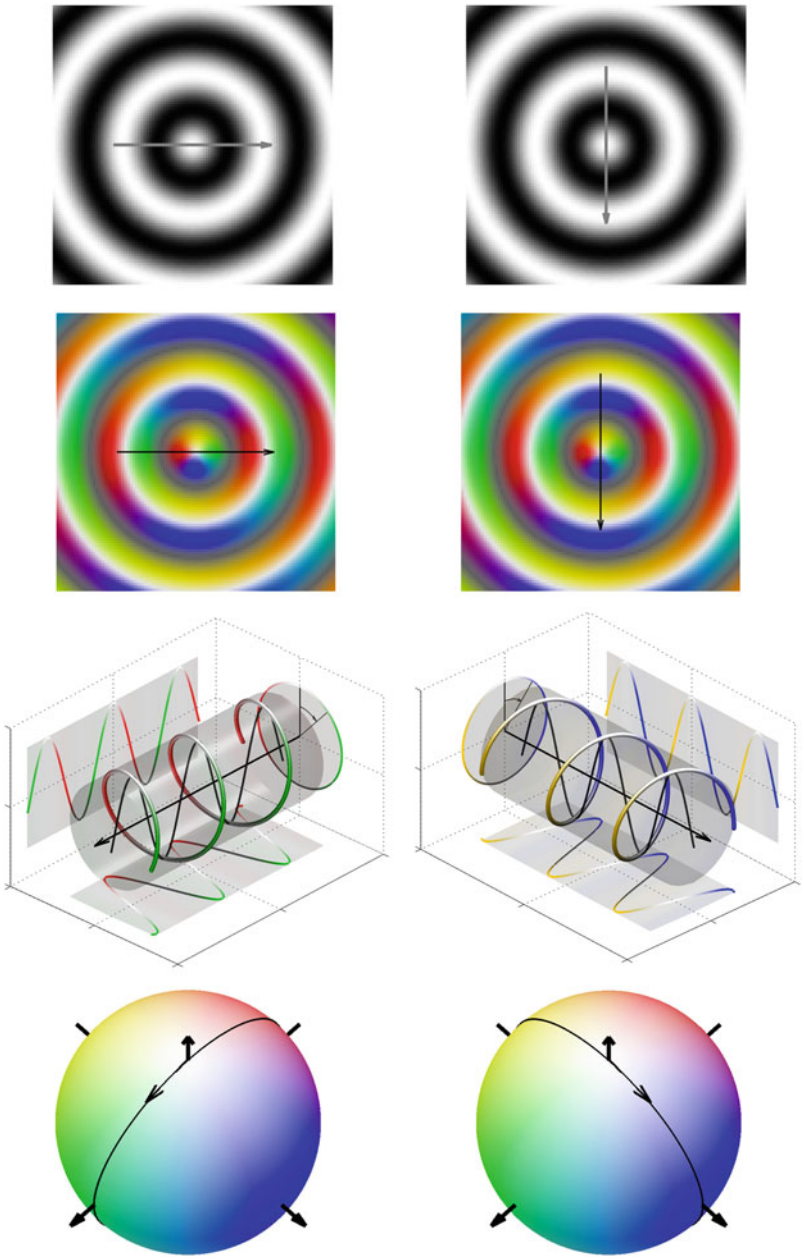


Fig. 9 Plots of the N + 1 dimensional vector phase space for 2-dimensional signals. *Left*: 3D coordinate system with direction angle, φ, and complex phase angle, θ. *Middle*: the same space color coded seen from above. *White* means $\theta = 0$, increased saturation means higher θ. Full saturation means $\theta = \pi/2$, the color indicates the direction, φ. A *black* iso-line is drawn at $\theta = \pi/4$. *Right*: the same color coded space seen from below. *Gray* means $\theta = \pi$. A *black* iso-line is drawn at $\theta = 3\pi/4$

2.6.1 One-Dimensional Case

For one-dimensional signals the monomial response matrix, $\mathbf{Q}_{01}$, reduces to the original definition of amplitude and phase for analytic signals, [7], i.e.

$$\mathbf{Q}_{01} = |\mathbf{Q}_{01}|\,\boldsymbol{\Psi}_{01} = Ae^{i\,\theta} \tag{35}$$

2.6.2 Monomial Structure Tensors

Although it is not the focus of the present chapter it seems appropriate to here mention that using the monomial filter response matrix the corresponding local structure tensor is, as outlined in [29], found as:

$$\mathbf{T}^2_{nm} = \mathbf{Q}_{nm}\,\mathbf{Q}^T_{nm} \tag{36}$$

2.6.3 Phase from Directed Quadrature Filters

The construction of classical quadrature filter pairs, [8,17], is a generalization of the analytic signal concept to higher dimensions but require pre-defined filter directions. A simple approach to local phase analysis of multidimensional signals is to measure a number of 1-dimensional phases from a set of quadrature filters given a number of fixed directions. The complex scalar response of a quadrature filter can, in the traditional manner, be expressed in terms of local amplitude, $\mathsf{A} = |\,\mathsf{q}_{\boldsymbol{k}}\,|$, and local phase, $\psi_{\boldsymbol{k}} = \hat{\mathsf{q}}_{\boldsymbol{k}}$ or phase angle, $\theta_{\boldsymbol{k}} = \arg(\mathsf{q}_{\boldsymbol{k}})$.

$$\mathsf{q}_{\boldsymbol{k}} \;=\; \mathsf{A}\;\psi_{\boldsymbol{k}} \;=\; \mathsf{A}\;e^{i\,\theta_{\boldsymbol{k}}} \tag{37}$$

A problem is that $\psi_{-k} = \psi_k^*$ implying that the same information can end up being represented in two ways, i.e. the representation is not one-to-one. For simple signals the measured phase will be invariant to the filter direction (except for the conjugate problem above) but for non-simple signals the directional dependence of the local phase can be highly complex and difficult to interpret. Although a lot of information about the local phase will be present in the joint outputs of a set of quadrature filters in different directions we do not consider it a preferable starting point for producing a good phase representation.

2.6.4 The N + 1 Dimensional Phase Vector

For 2-dimensional signals the monomial phase, $\boldsymbol{\Psi}_{01}$, is equivalent to the 3-dimensional vector phase definition for 2-dimensional signals due to Knutsson presented in [10] (1989). For higher dimensional signal spaces it generalizes to the N + 1 dimensional vector phase definition given in [17] (1995), i.e.

$$\Psi_{N+1} = \Psi_{01} = \frac{\mathbf{Q}_{01}}{\|\mathbf{Q}_{01}\|} = \frac{\{\, \mathsf{A}_0 \;,\; \mathsf{A}_1 \; \hat{\mathbf{u}} \,\}}{\sqrt{\mathsf{A}_0^2 + \mathsf{A}_1^2}} \tag{38}$$

2.6.5 Monogenic Signal

The Felsberg monogenic signal, F_M, defined for 2-dimensional signals, [18] (2001), will in our notation be written:

$$F_M = \mathbf{Q}_{01} = \{\, \mathsf{A}_0 \;,\; \mathsf{A}_1 \; \hat{\mathbf{u}} \,\} \tag{39}$$

In other words it is identical to the monomial filter response matrix using orders 0 and 1. An extension to higher dimensional signals is straight forward and will also be expressed by Eq. (39). This relates directly to the $N+1$ *dimensional phase vector* above and to the *conformal monogenic signal*, [28, 30].

2.6.6 Monogenic Phase Vector

The monogenic phase, φ_M, defined for 2-dimensional signals, [18] (2001), is found by the following mapping of the monogenic signal:

$$\varphi_M = \tan^{-1}\left(\frac{\|\,\mathsf{A}_1\,\|}{\mathsf{A}_0}\right)\hat{\mathbf{u}}_\perp \tag{40}$$

where $\tan^{-1} \in [\,0,\ \pi\,[$ and $\hat{\mathbf{u}}_\perp = (-\mathsf{u}_y, \mathsf{u}_x)^T$ if $\hat{\mathbf{u}} = (\mathsf{u}_x, \mathsf{u}_y)^T$. The definition monogenic phase is formulated in the 3-dimensional monogenic signal space and $\hat{\mathbf{u}}_\perp$ is obtained via the cross product $(1,0,0)^T \times (\mathsf{A}_0,\ \mathsf{A}_1\mathsf{u}_x,\ \mathsf{A}_1\mathsf{u}_y)^T$. Note that, since the definition is based on the cross product, it does not directly generalize to higher dimensions.

2.6.7 Three-Dimensional Case

For three-dimensional signal spaces $\boldsymbol{\Psi}_{01}$ is equivalent to a unitary quaternion.

3 Higher Order Phase Representations

The N + 1 dimensional phase representation and the monogenic signal definitely have important merits but, as we will show in this section, still leave some things to be desired. The monomial phase approach, however, allows for a higher order phase representation that will meet our requirements.

As a starting point for the discussions regarding higher order phase representations we state that all information except the magnitude should be represented by the phase. For a collection of monomial filter response matrices, $\{\mathbf{Q}_{n_i}\}$, let a generalized phase be defined by:

$$\boldsymbol{\Psi}_{\{n_i\}} = \{\hat{\mathbf{Q}}_{n_i}\} \equiv \frac{\{\mathbf{Q}_{n_i}\}}{\|\{\mathbf{Q}_{n_i}\}\|_{\mathsf{fro}}} \tag{41}$$

where 'fro' indicates the Frobenius norm (also known as the Hilbert-Schmidt norm) which is simply obtained as the square root of the sum of the absolute squares of all individual scalar elements.

3.1 Local Phase from Orders 0, 1, and 2

We can attain a joint representation of signal orientation and phase using Monomial filters of orders 0, 1 and 2. Consider the following set of filter responses:

$$\mathbf{Q}_{012} = \{\alpha \mathsf{Q}_0,\ \beta \mathbf{Q}_1,\ \gamma \mathbf{Q}_2\} \tag{42}$$

Separate amplitude A and phase $\boldsymbol{\Psi}$

$$\mathsf{A} = ||\{\alpha \mathsf{Q}_0,\ \beta \mathbf{Q}_1,\ \gamma \mathbf{Q}_2\}||_{\mathsf{fro}} \tag{43}$$

$$\boldsymbol{\Psi} = \frac{\{\alpha \mathsf{Q}_0,\ \beta \mathbf{Q}_1,\ \gamma \mathbf{Q}_2\}}{\mathsf{A}} \tag{44}$$

The filter response set can now be written

$$\mathbf{Q}_{012} = \mathsf{A}\boldsymbol{\Psi} \tag{45}$$

Note that in practise Q_0 need not be computed using a separate monomial filter of order 0 since an even/odd monomial filter of order n span all subspaces of lower even/odd order. e.g. for the present case: $\mathsf{Q}_0 = \text{trace}(\mathbf{Q}_2)$.

3.1.1 Balance Between Order 0, 1 and 2 Components

Equation (44) gives us the general form of a phase representation, however, we have yet to determine the relative weights, α, β and γ, of the individual components To do this it suffices to study the sinusoidal signal case. For a sinusoidal signal with orientation $\hat{\mathbf{u}}$ and phase θ we then get the following expression for the phase, $\boldsymbol{\Psi}$.

$$\boldsymbol{\Psi}(\theta, \hat{\mathbf{u}}) = \left\{\alpha \cos(\theta),\ i\,\beta \sin(\theta)\,\hat{\mathbf{u}}^T,\ \gamma \cos(\theta)\,\hat{\mathbf{u}}\,\hat{\mathbf{u}}^T\right\} \tag{46}$$

In order to find appropriate values for α, β and γ we need to add some constraints. We want the behavior of the representation to be as simple as possible and it is natural to require that the sensitivity of the phase representation to a change in local phase angle, as well as a change in local orientation, should be invariant to the local phase. In other words we will, as stated above, require that the metric is shift invariant, i.e. the norm of the partial derivatives with respect to θ and $\hat{\mathbf{u}}$ should be constants.

$$\begin{cases} \| \frac{\partial \boldsymbol{\Psi}}{\partial \theta} \|_{\text{fro}} = c_1 > 0 \\ \| \frac{\partial \boldsymbol{\Psi}}{\partial \hat{\mathbf{u}}} \|_{\text{fro}} = c_2 > 0 \end{cases} \tag{47}$$

Calculating the squared norm of $\frac{\partial \boldsymbol{\Psi}}{\partial \theta}$ gives:

$$\begin{aligned} \left\| \frac{\partial \boldsymbol{\Psi}}{\partial \theta} \right\|_{\text{fro}}^2 &= \left\| \left\{ -\alpha \sin(\theta),\ i\,\beta \cos(\theta)\, \hat{\mathbf{u}}^T,\ -\gamma \sin(\theta)\, \hat{\mathbf{u}}\, \hat{\mathbf{u}}^T \right\} \right\|_{\text{fro}}^2 \\ &= (\alpha^2 + \gamma^2) \sin^2(\theta) + \beta^2 \cos^2(\theta) \end{aligned} \tag{48}$$

By inspection it is simple to see that if we set

$$\beta^2 = \alpha^2 + \gamma^2 \tag{49}$$

we get:

$$\left\| \frac{\partial \boldsymbol{\Psi}}{\partial \theta} \right\|_{\text{fro}}^2 = \beta^2 \tag{50}$$

and the first requirement is met.

Finding the solution to the second requirement is slightly more involved. We will revert to the definition of partial derivatives by studying the change when adding an infinitesimal perpendicular vector $\epsilon\, \hat{\mathbf{v}}$ to $\hat{\mathbf{u}}$ with $\| \hat{\mathbf{v}} \| = 1$. Omitting the ϵ^2 terms, we get:

$$\begin{aligned} &\boldsymbol{\Psi}(\theta, \hat{\mathbf{u}} + \epsilon \hat{\mathbf{v}}) = \\ &= \left\{ \alpha \cos(\theta),\ i\beta \sin(\theta)(\hat{\mathbf{u}} + \epsilon \hat{\mathbf{v}})^T,\ \gamma \cos(\theta)(\hat{\mathbf{u}} + \epsilon \hat{\mathbf{v}})(\hat{\mathbf{u}} + \epsilon \hat{\mathbf{v}})^T \right\} \\ &= \left\{ \alpha \cos(\theta),\ i\beta \sin(\theta)(\hat{\mathbf{u}} + \epsilon \hat{\mathbf{v}})^T,\ \gamma \cos(\theta)[\hat{\mathbf{u}}\, \hat{\mathbf{u}}^T + \epsilon(\hat{\mathbf{v}}\, \hat{\mathbf{u}}^T + \hat{\mathbf{u}}\, \hat{\mathbf{v}}^T)] \right\} \end{aligned} \tag{51}$$

Expressing the difference in $\boldsymbol{\Psi}$ yields:

$$\boldsymbol{\Psi}(\theta, \hat{\mathbf{u}} + \epsilon \hat{\mathbf{v}}) - \boldsymbol{\Psi}(\theta, \hat{\mathbf{u}}) = \left\{ 0,\ i\beta \sin(\theta) \epsilon\, \hat{\mathbf{v}}^T,\ \gamma \cos(\theta) \epsilon (\hat{\mathbf{v}}\, \hat{\mathbf{u}}^T + \hat{\mathbf{u}}\, \hat{\mathbf{v}}^T) \right\} \tag{52}$$

Dividing by ϵ and finding the limit as $\epsilon \to 0$ we arrive at:

$$\begin{aligned}\frac{\partial \boldsymbol{\Psi}}{\partial\, \hat{\mathbf{u}}} &= \lim_{\epsilon \to 0} \frac{\boldsymbol{\Psi}(\theta,\, \hat{\mathbf{u}} + \epsilon\, \hat{\mathbf{v}}) - \boldsymbol{\Psi}(\theta,\, \hat{\mathbf{u}})}{\epsilon} \\ &= \left\{0,\ i\beta \sin(\theta)\, \hat{\mathbf{v}}^{\,T},\ \gamma \cos(\theta)(\, \hat{\mathbf{v}}\, \hat{\mathbf{u}}^{\,T} + \hat{\mathbf{u}}\, \hat{\mathbf{v}}^{\,T})\right\}\end{aligned} \tag{53}$$

Calculating the squared norm of $\frac{\partial \boldsymbol{\Psi}}{\partial\, \hat{\mathbf{u}}}$ gives:

$$\left\| \frac{\partial \boldsymbol{\Psi}}{\partial\, \hat{\mathbf{u}}} \right\|_{\mathsf{fro}}^2 = \beta^2 \sin^2(\theta) + \gamma^2 \cos^2 \| \, \hat{\mathbf{v}}\, \hat{\mathbf{u}}^{\,T} + \, \hat{\mathbf{u}}\, \hat{\mathbf{v}}^{\,T} \|_{\mathsf{fro}}^2 \tag{54}$$

and, as we now that $\hat{\mathbf{v}}^{\,T}\, \hat{\mathbf{u}} = 0$

$$\left\| \frac{\partial \boldsymbol{\Psi}}{\partial\, \hat{\mathbf{u}}} \right\|_{\mathsf{fro}}^2 = \beta^2 \sin^2(\theta) + 2\, \gamma^2 \cos^2(\theta) \tag{55}$$

Setting

$$\beta^2 = 2\gamma^2 \tag{56}$$

gives

$$\left\| \frac{\partial \boldsymbol{\Psi}}{\partial\, \hat{\mathbf{u}}} \right\|_{\mathsf{fro}}^2 = \beta^2 \tag{57}$$

and the second requirement is also met showing that this phase representation is shift invariant with respect to $(\, \hat{\mathbf{u}}\,, \theta)$ for all $(\, \hat{\mathbf{u}}\,, \theta)$. A shift invariant metric trivially implies that the mapping is continuous. It is also not difficult to see that $\boldsymbol{\Psi}$ is unique with respect to $(\, \hat{\mathbf{u}}\,, \theta)$ and the exercise is left to the devoted reader.

At this stage the only degree of freedom left relates directly to the norm of the representation. The choice will not effect the general behavior of the representation and choosing

$$\alpha^2 + \beta^2 + \gamma^2 = 4 \tag{58}$$

gives the following particularly simple and unique solution:

$$\alpha = 1, \quad \beta = \sqrt{2}, \quad \gamma = 1 \tag{59}$$

With this choice the phase representation is composed of the following components:

$$\boldsymbol{\Psi}(\theta,\, \hat{\mathbf{u}}) = \left\{\cos(\theta),\ i\, \sqrt{2} \sin(\theta)\, \hat{\mathbf{u}}^{\,T},\ \cos(\theta)\, \hat{\mathbf{u}}\, \hat{\mathbf{u}}^{\,T} \right\} \tag{60}$$

3.1.2 A Comment on Some Alternative Phase Representations

The investigation above shows that all three orders are needed to construct a phase representation that meets the shift invariance requirement. This disqualifies, for example, the N + 1 dimensional Haglund-Knutsson-Granlund local phase (1989), [10], and the Felsberg-Sommer monogenic signal (2001), [18], both of which only use orders 0 and 1, and are in fact identical except for an amplitude normalisation. The monogenic phase is a mapping of the monogenic signal and can be viewed as an attempt to construct a concept that is more similar to the traditional phase angle in one-dimensional signal analysis. A consequence of the similarity with the traditional phase angle is that the 'wrap around' discontinuity is also inherited. This will not be a problem if the monogenic phase estimate is considered the final result of the analysis, However, if the estimates are used as input to further processing this mapping of the monogenic signal will make it difficult to apply many standard operations, e.g. averaging and differentiation, due to the discontinuous representation.

3.2 A Matrix Representation of Phase

A matrix phase representation that meets all the above requirements is given by:

$$\boldsymbol{\Psi}(\theta,\ \hat{\mathbf{u}}) \ = \ \begin{pmatrix} \mathsf{u} & i\,\mathbf{u}^T \\ i\,\mathbf{u} & \mathbf{U} \end{pmatrix} \ = \ \begin{pmatrix} \cos(\theta) & i\,\sin(\theta)\ \hat{\mathbf{u}}^T \\ i\,\sin(\theta)\ \hat{\mathbf{u}} & \cos(\theta)\ \hat{\mathbf{u}}\,\hat{\mathbf{u}}^T \end{pmatrix} \tag{61}$$

A number of interesting observations regarding this matrix representation can be made. The representation has a Frobenius norm of $\sqrt{2}$ and two non-zero eigenvalues with unity norm. The eigenvalues are a complex conjugate pair implying an order 2 sub-determinant of unity. The representation is pseudo-unitary, i.e. unitary in the spanned 2-dimensional subspace, implying:

$$\begin{cases} \boldsymbol{\Psi}(\theta,\ \hat{\mathbf{u}})^\dagger = \boldsymbol{\Psi}(\theta,\ \hat{\mathbf{u}})^+ \\ \boldsymbol{\Psi}(\theta,\ \hat{\mathbf{u}})^\dagger\ \boldsymbol{\Psi}(\theta,\ \hat{\mathbf{u}}) = \boldsymbol{\Psi}(\theta,\ \hat{\mathbf{u}})\ \boldsymbol{\Psi}(\theta,\ \hat{\mathbf{u}})^\dagger = \mathbf{P}_2 \end{cases} \tag{62}$$

where '†' denotes the complex matrix transpose, '$^+$' denotes the Moore-Penrose pseudo inverse and $\mathbf{P}_2$ is a projection matrix having two non-zero eigenvalues equal to unity.

3.3 The Phase Space Manifold Is a Klein Bottle

By plotting image patches with different phase and orientation in a regular fashion the intrinsic Klein bottle structure of the phase space manifold is revealed, [32], see Fig. 10 and the caption. A Klein bottle, K, is a non-orientable surface in four

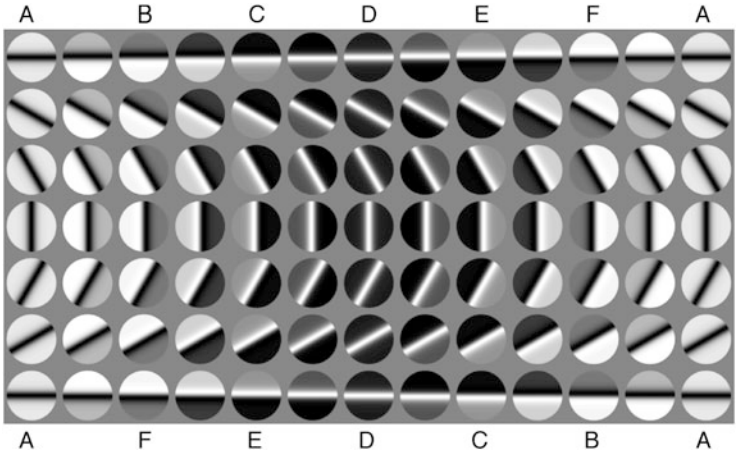


Fig. 10 By plotting image patches with different phase and orientation in a regular fashion the intrinsic Klein bottle structure of the phase space is revealed. Orientation changes from top to bottom and phase changes from left to right. The first and last row and the first and last column represent the same orientation/phase combinations. To match the first and last row one of them has to be taken in reversed order. This shows that, for 2-dimensional signals, the phase manifold is a Klein bottle

dimensions and for 2-dimensional simple signals the phase matrix, Ψ, only needs four degrees of freedom. In general, however, the phase matrix, Ψ, has five degrees of freedom. The fifth dimension is needed to represent the local phase of non-simple signals. That is the general topology of the phase matrix for two-dimensional signals is $K \times I$, where I is an interval of real numbers.

3.3.1 Color Code

The five degrees of freedom of the phase matrix, Ψ, are hard to directly visualize in a simple way. To obtain images that holds as much as possible of the phase information we will introduce a color coding scheme. This implies a mapping from five dimensions to the three dimensions of the color space used in Fig. 8. The hue is based on the direction of the imaginary vector part of the phase matrix product, i.e. phase matrix elements $\Psi(2,1)$ and $\Psi(3,1)$. Zero-phase (ridge/white line) is coded as white and 180 phase (valley/black line) as gray. See Figs. 11 and 12.

4 Monomial Phase Products and Phase Angle Differences

An important operation using the complex scalar phase is the complex conjugate product giving a phase angle difference estimate. Let us see if we can use the new high order phase in the same way.

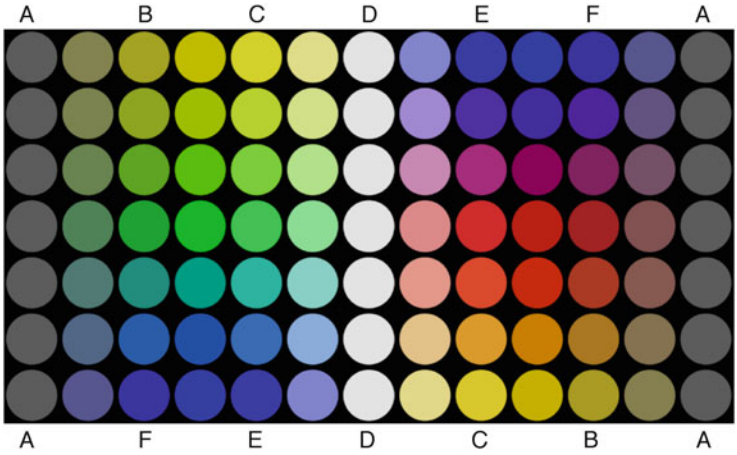


Fig. 11 Plot showing the colors corresponding to the phase at the center of the patches in Fig. 10. The figure provides a link between Figs. 10 and 12. It also shows that the RGB color space is inadequate to represent the Klein bottle topology since all the $\theta = 0$ and all the $\theta = \pi$ patches are *white* and *gray* respectively

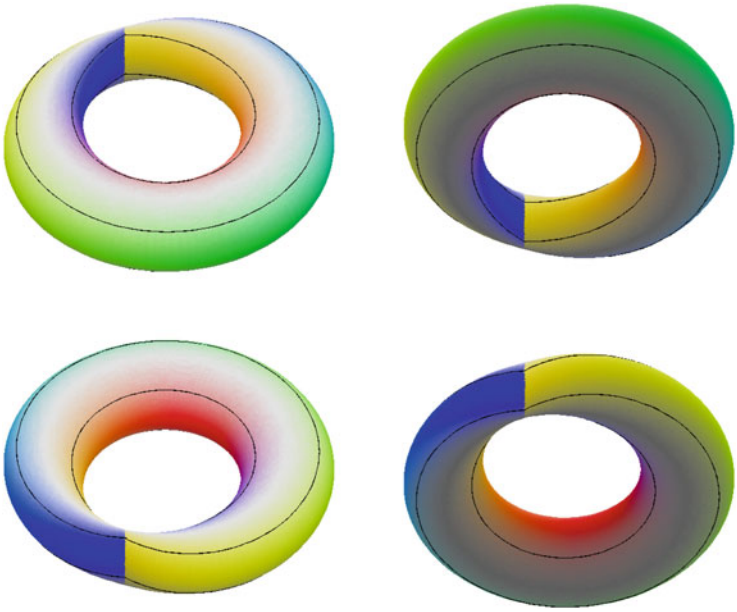


Fig. 12 Figure showing four views of a three-dimensional sub-space of the phase matrix representation. The Klein bottle structure shows itself like a modified torus (in a way similar to a mobius-band), a tube bent in a *circle* but before 'glued' together inside and outside surfaces trade places. The *black lines* indicate $\theta = \pm\pi/4$

4.1 Direct Products of Phase Matrices

In order to do this we need to decide how an appropriate product of two phase matrices can be defined. Starting with two phase matrices, $\boldsymbol{\Psi}_a$ and $\boldsymbol{\Psi}_b$:

$$\boldsymbol{\Psi}_a = \begin{pmatrix} \mathsf{a} & i\,\mathbf{a}^T \\ i\,\mathbf{a} & \mathbf{A} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & i\sin(\theta)\,\hat{\mathbf{a}}^T \\ i\sin(\theta)\,\hat{\mathbf{a}} & \cos(\theta)\,\hat{\mathbf{a}}\,\hat{\mathbf{a}}^T \end{pmatrix} \tag{63}$$

$$\boldsymbol{\Psi}_b = \begin{pmatrix} \mathsf{b} & i\,\mathbf{b}^T \\ i\,\mathbf{b} & \mathbf{B} \end{pmatrix} = \begin{pmatrix} \cos(\phi) & i\sin(\phi)\,\hat{\mathbf{b}}^T \\ i\sin(\phi)\,\hat{\mathbf{b}} & \cos(\phi)\,\hat{\mathbf{b}}\,\hat{\mathbf{b}}^T \end{pmatrix} \tag{64}$$

we will, in analogy with standard matrix multiplication, define the product of two phase matrices to be:

$$\boldsymbol{\Psi}_a\,\boldsymbol{\Psi}_b = \begin{pmatrix} \mathsf{ab} - \mathbf{a}^T\mathbf{b} & i\,(\mathsf{a}\mathbf{b}^T + \mathbf{a}^T\mathbf{B}) \\ i\,(\mathsf{a}\mathbf{b} + \mathbf{A}\mathbf{b}) & -\mathbf{a}\mathbf{b}^T + \mathbf{A}\mathbf{B} \end{pmatrix} \tag{65}$$

Note that the components multiplied here can be scalars, vectors or matrices. All products are, however, standard matrix algebra products.

As mentioned above we wish to investigate the analog of the complex conjugate product for complex numbers. For two phase matrices this translates to the complex transpose product, i.e. terms holding $\mathbf{b}$ will change sign. Letting ' † ' denote the complex conjugate transpose we get:

$$\begin{aligned} \boldsymbol{\Psi}_a\,\boldsymbol{\Psi}_b^\dagger &= \begin{pmatrix} \mathsf{ab} + \mathbf{a}^T\mathbf{b} & i\,(-\mathsf{a}\mathbf{b}^T + \mathbf{a}^T\mathbf{B}) \\ i\,(\mathsf{a}\mathbf{b} - \mathbf{A}\mathbf{b}) & \mathbf{a}\mathbf{b}^T + \mathbf{A}\mathbf{B} \end{pmatrix} \\ &= \begin{pmatrix} \mathsf{c}_{11} & i\,\mathbf{c}_{12}^T \\ i\,\mathbf{c}_{21} & \mathbf{C}_{22} \end{pmatrix} \end{aligned} \tag{66}$$

The last entry in Eq. (66) simply serves as a way to identify the individual components of $\boldsymbol{\Psi}_a\,\boldsymbol{\Psi}_b^\dagger$. Carrying out the calculations for each component we obtain:

$$\begin{aligned} \mathsf{c}_{11} &= \cos(\theta)\cos(\phi) + \sin(\theta)\sin(\phi)\,\hat{\mathbf{a}}^T\hat{\mathbf{b}} \\ \mathbf{c}_{12}^T &= -\cos(\theta)\sin(\phi)\,\hat{\mathbf{b}}^T + \sin(\theta)\cos(\phi)\,\hat{\mathbf{a}}^T\hat{\mathbf{b}}\,\hat{\mathbf{b}}^T \\ \mathbf{c}_{21} &= \sin(\theta)\cos(\phi)\,\hat{\mathbf{a}} - \cos(\theta)\sin(\phi)\,\hat{\mathbf{a}}\,\hat{\mathbf{a}}^T\hat{\mathbf{b}} \\ \mathbf{C}_{22} &= \sin(\theta)\sin(\phi)\,\hat{\mathbf{a}}\,\hat{\mathbf{b}}^T + \cos(\theta)\cos(\phi)\,\hat{\mathbf{a}}\,\hat{\mathbf{a}}^T\hat{\mathbf{b}}\,\hat{\mathbf{b}}^T \end{aligned} \tag{67}$$

Thus the result in the general case is quite complex, however, setting the orientations of the to phase matrices to be equal will allow us to significantly simplify the expression.

4.1.1 Equal Orientations

For the equal orientation case, i.e. $\hat{\mathbf{b}} = \hat{\mathbf{a}}$, we get:

$$\begin{aligned}
\mathsf{c}_{11} &= \cos(\theta)\cos(\phi) + \sin(\theta)\sin(\phi) \\
\mathbf{c}_{12}^T &= [-\cos(\theta)\sin(\phi) + \sin(\theta)\cos(\phi)]\ \hat{\mathbf{a}}^T \\
\mathbf{c}_{21} &= [\ \sin(\theta)\cos(\phi) - \cos(\theta)\sin(\phi)]\ \hat{\mathbf{a}} \\
\mathbf{C}_{22} &= [\ \sin(\theta)\sin(\phi) + \cos(\theta)\cos(\phi)]\ \hat{\mathbf{a}}\,\hat{\mathbf{a}}^T
\end{aligned} \tag{68}$$

Simplifying further and restoring the matrix notation we arrive at the following expression:

$$\begin{aligned}
\boldsymbol{\Psi}_a\,\boldsymbol{\Psi}_b^{\dagger} &= \begin{pmatrix} \cos(\theta-\phi) & i\ \sin(\theta-\phi)\ \hat{\mathbf{a}}^T \\ i\ \sin(\theta-\phi)\ \hat{\mathbf{a}} & \cos(\theta-\phi)\ \hat{\mathbf{a}}\,\hat{\mathbf{a}}^T \end{pmatrix} \\
&= \boldsymbol{\Psi}(\,\theta-\phi,\ \hat{\mathbf{a}}\,) \\
&= \boldsymbol{\Psi}(\,\Delta\theta,\ \hat{\mathbf{a}}\,)
\end{aligned} \tag{69}$$

where $\Delta\theta$ is the difference in phase angle. In other words, it works perfectly for the "same orientation" case. We get a direct analogue to the multiplication of complex numbers.

4.2 A Symmetric Complex Conjugate Phase Product

The simple matrix product analogue defined in Eq. (65) has a significant drawback. The result will not have the same symmetry features as the multiplied phase matrices unless the orientations are identical. For this reason we define a symmetric complex conjugate phase matrix product that will retain the symmetry features also when the orientations differ.

$$\begin{aligned}
\Psi_{\frac{a}{b}} &\stackrel{\text{def}}{=} \frac{1}{2}(\boldsymbol{\Psi}_a\,\boldsymbol{\Psi}_b^{\dagger} + \boldsymbol{\Psi}_b^{\dagger}\,\boldsymbol{\Psi}_a) \\
&= \frac{1}{2}\begin{pmatrix} \mathsf{ab} + \mathbf{a}^T\mathbf{b} & i\,(-\mathsf{ab}^T + \mathbf{a}^T\mathbf{B}) \\ i\,(\mathsf{ab} - \mathbf{A}\mathsf{b}) & \mathsf{ab}^T + \mathbf{AB} \end{pmatrix} \\
&+ \frac{1}{2}\begin{pmatrix} \mathsf{ba} + \mathbf{b}^T\mathbf{a} & i\,(\ \mathsf{ba}^T - \mathbf{b}^T\mathbf{A}) \\ i\,(-\mathsf{ba} + \mathbf{B}\mathsf{a}) & \mathsf{ba}^T + \mathbf{BA} \end{pmatrix} \\
&= \begin{pmatrix} \mathsf{c}_{11} & i\,\mathbf{c}_{12}^T \\ i\,\mathbf{c}_{21} & \mathbf{C}_{22} \end{pmatrix}
\end{aligned} \tag{70}$$

Carrying out the calculations for each component we get:

$$\begin{aligned}
\mathbf{c}_{11} &= \tfrac{1}{2}\,(\ \cos(\theta)\cos(\phi) \quad + \ \sin(\theta)\sin(\phi)\ \hat{\mathbf{a}}^T\hat{\mathbf{b}} \\
&\quad + \cos(\theta)\cos(\phi) \quad + \ \sin(\theta)\sin(\phi)\ \hat{\mathbf{b}}^T\hat{\mathbf{a}}\ ) \\
\mathbf{c}_{12}^T &= \tfrac{1}{2}\,(-\cos(\theta)\sin(\phi)\ \hat{\mathbf{b}}^T + \ \sin(\theta)\cos(\phi)\ \hat{\mathbf{a}}^T\hat{\mathbf{b}}\,\hat{\mathbf{b}}^T \\
&\quad + \cos(\phi)\sin(\theta)\ \hat{\mathbf{a}}^T - \ \sin(\phi)\cos(\theta)\ \hat{\mathbf{b}}^T\hat{\mathbf{a}}\,\hat{\mathbf{a}}^T\ ) \\
\mathbf{c}_{21} &= \tfrac{1}{2}\,(\ \sin(\theta)\cos(\phi)\ \hat{\mathbf{a}} \quad - \ \cos(\theta)\sin(\phi)\ \hat{\mathbf{a}}\,\hat{\mathbf{a}}^T\hat{\mathbf{b}} \\
&\quad - \sin(\phi)\cos(\theta)\ \hat{\mathbf{b}} \quad + \ \cos(\phi)\sin(\theta)\ \hat{\mathbf{b}}\,\hat{\mathbf{b}}^T\hat{\mathbf{a}}\ ) \\
\mathbf{C}_{22} &= \tfrac{1}{2}(\ \ \sin(\theta)\sin(\phi)\ \hat{\mathbf{a}}\,\hat{\mathbf{b}}^T + \ \cos(\theta)\cos(\phi)\ \hat{\mathbf{a}}\,\hat{\mathbf{a}}^T\hat{\mathbf{b}}\,\hat{\mathbf{b}}^T \\
&\quad + \sin(\theta)\sin(\phi)\ \hat{\mathbf{b}}\,\hat{\mathbf{a}}^T + \ \cos(\theta)\cos(\phi)\ \hat{\mathbf{b}}\,\hat{\mathbf{b}}^T\hat{\mathbf{a}}\,\hat{\mathbf{a}}^T)
\end{aligned} \tag{71}$$

Setting $\eta = \hat{\mathbf{a}}^T\hat{\mathbf{b}}$ and simplifying gives:

$$\begin{aligned}
\mathbf{c}_{11} &= \quad \cos(\theta)\cos(\phi) \ + \ \eta\sin(\theta)\sin(\phi) \\
\mathbf{c}_{21} &= \tfrac{1}{2}\,[\ \ (\sin(\theta)\cos(\phi) \ - \ \eta\cos(\theta)\sin(\phi))\ \hat{\mathbf{a}} \\
&\quad +(-\sin(\phi)\cos(\theta) \ + \ \eta\cos(\phi)\sin(\theta)\ )\ \hat{\mathbf{b}}\ \ ] \\
\mathbf{c}_{12} &= \mathbf{c}_{21} \\
\mathbf{C}_{22} &= \tfrac{1}{2}(\ \ \sin(\theta)\sin(\phi) \ + \ \eta\cos(\theta)\cos(\phi)\ )(\hat{\mathbf{a}}\,\hat{\mathbf{b}}^T + \ \hat{\mathbf{b}}\,\hat{\mathbf{a}}^T)
\end{aligned} \tag{72}$$

Expressing using factors $1 \pm \eta$ we get:

$$\begin{aligned}
\mathbf{c}_{11} &= \tfrac{1}{2}[\ (1+\eta)(\cos(\theta)\cos(\phi) + \sin(\theta)\sin(\phi)) \\
&\quad + (1-\eta)(\cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi))\ ] \\
\mathbf{c}_{21} &= \tfrac{1}{4}[\ [(1+\eta)(\sin(\theta)\cos(\phi) - \cos(\theta)\sin(\phi)) \\
&\quad + (1-\eta)(\sin(\theta)\ \cos(\phi) + \cos(\theta)\sin(\phi))\ ]\ \hat{\mathbf{a}} \\
&\quad - [(1+\eta)(\sin(\phi)\cos(\theta) - \cos(\phi)\sin(\theta)) \\
&\quad + (1-\eta)(\sin(\phi)\cos(\theta) + \cos(\phi)\sin(\theta))\ ]\ \hat{\mathbf{b}}\ ] \\
\mathbf{c}_{12} &= \mathbf{c}_{21} \\
\mathbf{C}_{22} &= \tfrac{1}{4}[\ [\ (1+\eta)(\sin(\theta)\sin(\phi) + \cos(\theta)\cos(\phi)) \\
&\quad + (1-\eta)(\sin(\theta)\sin(\phi) - \cos(\theta)\cos(\phi))\ ](\hat{\mathbf{a}}\,\hat{\mathbf{b}}^T + \ \hat{\mathbf{b}}\,\hat{\mathbf{a}}^T)\ ]
\end{aligned} \tag{73}$$

Using classical trigonometrics we arrive at:

$$\begin{aligned}
\mathbf{c}_{11} &= \frac{1}{2}[\ (1+\eta)\cos(\theta-\phi)\ +\ (1-\eta)\cos(\theta+\phi)\,] \\
\mathbf{c}_{21} &= \frac{1}{4}[\,(1+\eta)\sin(\theta-\phi)(\hat{\mathbf{a}}\ +\ \hat{\mathbf{b}}\,)\ +(1-\eta)\sin(\theta+\phi)(\hat{\mathbf{a}}\ -\ \hat{\mathbf{b}}\,)\,] \\
\mathbf{c}_{12} &= \mathbf{c}_{21} \\
\mathbf{C}_{22} &= \frac{1}{4}[\,(1+\eta)\cos(\theta-\phi)\ -\ (1-\eta)\cos(\theta+\phi)\,]\,(\,\hat{\mathbf{a}}\,\hat{\mathbf{b}}^{\,T}+\,\hat{\mathbf{b}}\,\hat{\mathbf{a}}^{\,T})
\end{aligned} \tag{74}$$

In a final desperate attempt to understand what's going on we introduce the variables:

ζ = the angle between the two signal directions $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$
$\mathbf{s} = \frac{1}{2}(\,\hat{\mathbf{a}}\ +\ \hat{\mathbf{b}}\,)$ the mean of the two signal direction vectors
$\mathbf{d} = \frac{1}{2}(\,\hat{\mathbf{a}}\ -\ \hat{\mathbf{b}}\,)$ half of the difference between the two signal direction vectors.

Substituting the new variables and using standard trigonometric identities we find:

$$\begin{aligned}
\mathbf{c}_{11} &= \ \cos^2(\frac{\zeta}{2})\,\cos(\theta-\phi)\ \ +\ \ \sin^2(\frac{\zeta}{2})\,\cos(\theta+\phi) \\
\mathbf{c}_{21} &= \ [\cos^2(\frac{\zeta}{2})\,\sin(\theta-\phi)\,\mathbf{s}\ +\ \ \sin^2(\frac{\zeta}{2})\,\sin(\theta+\phi)\,\mathbf{d}\,] \\
\mathbf{c}_{12} &= \mathbf{c}_{21} \\
\mathbf{C}_{22} &= \ [\cos^2(\frac{\zeta}{2})\,\cos(\theta-\phi)\ \ -\ \ \sin^2(\frac{\zeta}{2})\,\cos(\theta+\phi)\,]\,(\mathbf{s}\,\mathbf{s}^{\,T}-\,\mathbf{d}\mathbf{d}^{\,T})
\end{aligned} \tag{75}$$

Noting that $\|\mathbf{s}\| = \cos(\frac{\zeta}{2})$ and $\|\mathbf{d}\| = \sin(\frac{\zeta}{2})$ gives:

$$\begin{cases}
\mathbf{s} = \cos(\frac{\zeta}{2})\,\hat{\mathbf{s}} \\
\mathbf{d} = \sin(\frac{\zeta}{2})\,\hat{\mathbf{d}}
\end{cases} \tag{76}$$

We can now write:

$$\begin{aligned}
\mathbf{c}_{11} &= \ \cos^2(\tfrac{\zeta}{2})\,\cos(\theta-\phi)\ \ +\ \ \sin^2(\tfrac{\zeta}{2})\,\cos(\theta+\phi) \\
\mathbf{c}_{21} &= \ [\cos^3(\tfrac{\zeta}{2})\,\sin(\theta-\phi)\,\hat{\mathbf{s}}\ +\ \ \sin^3(\tfrac{\zeta}{2})\,\sin(\theta+\phi)\,\hat{\mathbf{d}}\,] \\
\mathbf{c}_{12} &= \mathbf{c}_{21} \\
\mathbf{C}_{22} &= \ [\cos^4(\tfrac{\zeta}{2})\,\cos(\theta-\phi)\ \ -\tfrac{1}{4}\sin^2(\zeta)\,\cos(\theta+\phi)\,]\,\hat{\mathbf{s}}\,\hat{\mathbf{s}}^{\,T} \\
&\quad -[\,\tfrac{1}{4}\sin^2(\zeta)\,\cos(\theta-\phi)\ \ -\ \ \sin^4(\tfrac{\zeta}{2})\,\cos(\theta+\phi)\,]\,\hat{\mathbf{d}}\,\hat{\mathbf{d}}^{\,T}
\end{aligned} \tag{77}$$

Rearranging terms we can finally express $\boldsymbol{\Psi}_{\frac{a}{b}}$ in a way that will shed some light on what happens when $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ differs:

$$
\begin{aligned}
\boldsymbol{\Psi}_{\frac{a}{b}} \;=\; & \cos^2(\tfrac{\zeta}{2})\, \boldsymbol{\Psi}(\,\theta-\phi,\ \hat{\mathbf{s}}\,) \\
& -\cos^2(\tfrac{\zeta}{2}) \begin{pmatrix} 0 & i\,[1-\cos(\tfrac{\zeta}{2})]\ \sin(\theta-\phi)\ \hat{\mathbf{s}}^{\,T} \\ i\,[1-\cos(\tfrac{\zeta}{2})]\ \sin(\theta-\phi)\,\hat{\mathbf{s}} & [1-\cos^2(\tfrac{\zeta}{2})]\cos(\theta-\phi)\,\hat{\mathbf{s}}\,\hat{\mathbf{s}}^{\,T} \end{pmatrix} \\
& +\sin^2(\tfrac{\zeta}{2}) \begin{pmatrix} \cos(\theta+\phi) & i\ \sin(\tfrac{\zeta}{2})\ \sin(\theta+\phi)\ \hat{\mathbf{d}}^{\,T} \\ i\ \sin(\tfrac{\zeta}{2})\ \sin(\theta+\phi)\,\hat{\mathbf{d}} & \sin^2(\tfrac{\zeta}{2})\cos(\theta+\phi)\,\hat{\mathbf{d}}\,\hat{\mathbf{d}}^{\,T} \end{pmatrix} \\
& -\tfrac{1}{4}\sin^2(\zeta) \begin{pmatrix} 0 & \hat{\mathbf{0}}^{\,T} \\ \hat{\mathbf{0}} & \cos(\theta+\phi)\,\hat{\mathbf{s}}\,\hat{\mathbf{s}}^{\,T} \end{pmatrix} \\
& -\tfrac{1}{4}\sin^2(\zeta) \begin{pmatrix} 0 & \hat{\mathbf{0}}^{\,T} \\ \hat{\mathbf{0}} & \cos(\theta-\phi)\,\hat{\mathbf{d}}\,\hat{\mathbf{d}}^{\,T} \end{pmatrix}
\end{aligned}
\tag{78}
$$

We can see that the first term has the same form as the monomial phase matrix representation, except for a scaling with $\cos^2(\frac{\zeta}{2})$, and represents the difference in phase in the mean direction of the two signals. The magnitude of the four additional terms can be roughly described as follows. For the same orientation case, $\zeta = 0$, the additional terms are zero and otherwise no individual additional term has a larger norm than $\sin^2(\frac{\zeta}{2})$ and Eq. (78) can be written:

$$
\boldsymbol{\Psi}_{\frac{a}{b}} \;=\; \cos^2(\frac{\zeta}{2})\, \boldsymbol{\Psi}(\,\Delta\theta,\ \hat{\mathbf{s}}\,) \;+\; \sin^2(\frac{\zeta}{2})\boldsymbol{\Psi}_{\epsilon} \tag{79}
$$

where $\Delta\theta$ is the difference in phase angle, $(\theta - \phi)$, and $\boldsymbol{\Psi}_{\epsilon}$ is a matrix with terms having a limited magnitude. In other words, as long as the angular difference between the signals is not too large the result is very well-behaved. The experiments in the next section will demonstrate this behavior.

Two comments are appropriate to make here: (1) As is also the case for complex conjugate products of complex numbers, changing the order of the matrices (or complex numbers) will result in a complex transpose of the result, i.e.

$$
\Psi_{\frac{a}{b}} = \Psi^{\dagger}_{\frac{b}{a}} \tag{80}
$$

This simply corresponds to the fact that the change of order implies that the reference direction of the motion is reversed. (2) The symmetric complex conjugate

phase product is not associative, i.e. if carrying out products of more than two matrices changing the order in which the dyadic products are done may change the result. However, as the estimation always only involves two matrices, this has no consequence in the motion estimation case.

4.3 Monomial Phase and Motion Estimation

Local phase estimation is a key part of many displacement estimation algorithms. In this section we will demonstrate the power of the new phase representation in two different cases. We will refer to the computed phase based measure as the displacement even though the actual displacement in image coordinates is related by a factor dependent on the wavelength of the sine waves used. In all experiments the color coding used for the phase matrix component images on the left hand side is: *Green – positive real, Red – negative real, Blue – positive imaginary, Yellow – negative imaginary.* The Klein bottle sub-space color coding used for the images on the right hand side is described in Sect. 3.3. The displacement estimate used for the arrows is computed as follows. Expressing the symmetric complex conjugate matrix product, $\boldsymbol{\Psi}_{\frac{a}{b}}$, as:

$$\boldsymbol{\Psi}_{\frac{a}{b}} \;=\; \begin{pmatrix} \mathsf{c} & i\,\mathbf{c}^T \\ i\,\mathbf{c} & \mathbf{C} \end{pmatrix} \tag{81}$$

the simplest way to compute a displacement vector, $\mathbf{d}_\Psi$ is:

$$\mathbf{d}_\Psi = \tan^{-1}\left(\frac{\|\,\mathbf{c}\,\|}{\mathsf{c}}\right)\hat{\mathbf{c}} \;=\; \Delta\theta\;\hat{\mathbf{c}} \tag{82}$$

where $\tan^{-1} \in [0,\ \pi[$. This means that the direction of the displacement is given by $\hat{\mathbf{c}}$ and the displacement distance is computed as the arc-tangent of the ratio between the magnitude of $\mathbf{c}$ and the value of c. It is worth noting that the form of Eq. (82) has obvious similarities to the definition of the monogenic phase vector, Eq. (40). However, there are two important differences: Firstly, since the entries here are components of a matrix phase product, the result represents a phase relation rather than a single phase. Secondly, the formulation in Eq. (82) is valid for any signal dimensionality. It is also appropriate to note that in three dimensions Eq. (82) is equivalent to the logarithm of a unitary quaternion, $(\mathsf{q},\ \mathbf{q})$, i.e. $\tan^{-1}\left(\frac{\|\,\mathbf{q}\,\|}{\mathsf{q}}\right)\hat{\mathbf{q}} \;= \log(\mathsf{q},\ \mathbf{q})$.

4.3.1 Two Radial Sine Waves with Different Frequency

The simplest case, shown in Fig. 13 is a pair of radial sine wave patterns with slightly different frequency. The local orientation for this pair is identical and the phase difference grows linearly out from the center. The result in Fig. 13 is as predicted by the theoretical analysis, Eqs. (69), (81) and (82).

4.3.2 Two Sine Waves with Different Orientations

Figure 14 shows an example with two linear sine wave patterns with 20° difference in orientation. The result is as predicted by the theoretical analysis, Eqs. (79)–(82). This example clearly demonstrates the robustness of the approach.

4.3.3 A Comment on Optical Flow

Motion estimation and phase difference estimation are not identical problems but are related through a distance/phase relation. However, a comparison with the well-known optical flow motion estimation algorithm will highlight some important aspects. Figure 15 shows the result of the classical optical flow algorithm applied to the image pair in Fig. 14. As the gradient in optical flow is taken from a single reference image there are two possible solutions, shown left and right. The two fundamental assumptions made in optical flow are: (1) 'the image is locally planar' and (2) 'the second image is a translated version of the first image'. These to assumptions are in practice rarely a good approximation and are, in fact, severely violated for a simple translated sine wave pattern. As can be expected, using the gray images in Fig. 14 yield optical flow estimates that are quite useless. Although the conjugate phase products results in Fig. 14 are not equivalent to displacement estimates they demonstrate that the monomial phase based model is likely to provide a much more powerful approach in non-trivial cases.

4.4 Averaging of Phase Matrix Products

In many applications the presence of noise can render individual local feature estimates unreliable. A simple, powerful and extensively used method for noise suppression and regularization is low pass filtering of the obtained local estimates. However, and we stress this again, for the averaging operation to be meaningful is required that the feature representation space has the fundamental properties previously discussed in this chapter.

The appropriate spatio-temporal size of the low pass filter is determined by the noise level and how much 'smoothing' of the feature field that is acceptable. Here the importance of the symmetric complex conjugate phase product defined in Sect. 4

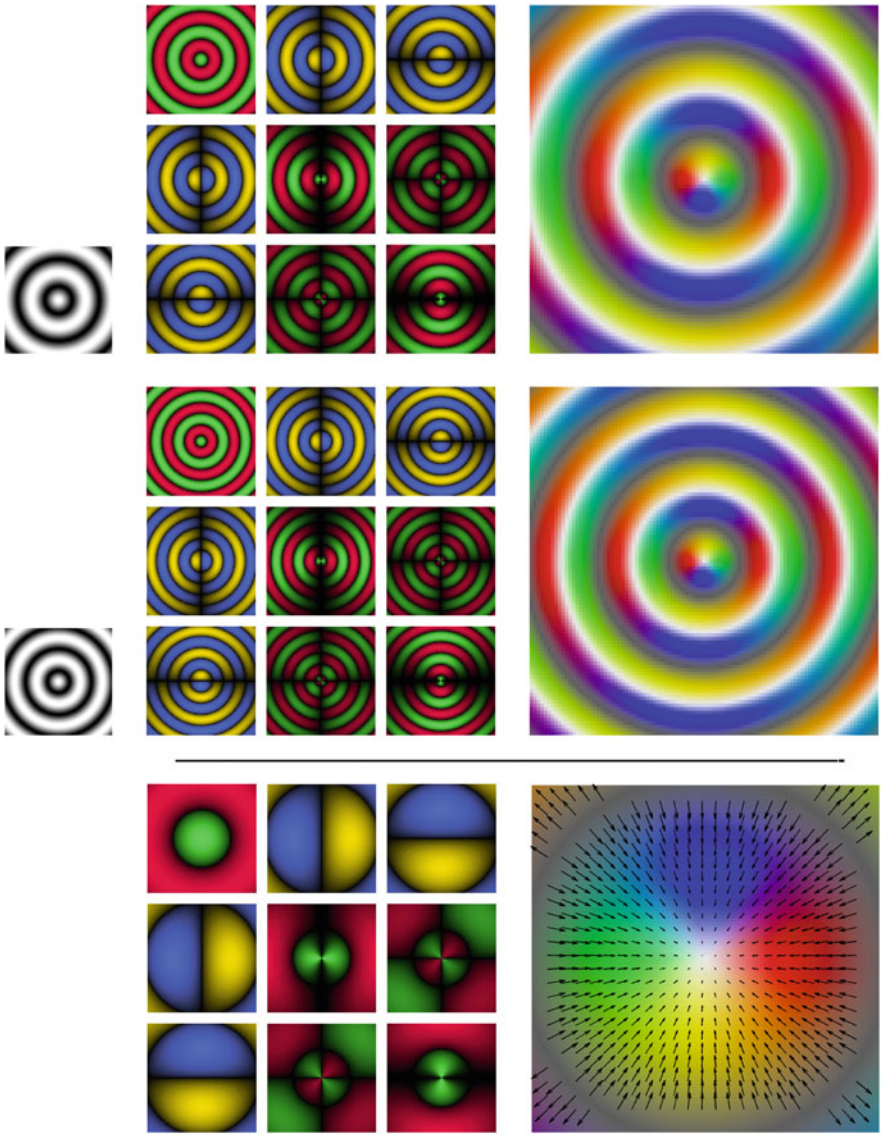


Fig. 13 Figure showing a simple scaling example using two radial sine waves with slightly different frequency. The *gray scale* images show the to different input images. The local orientation is identical in both images but the phase difference grows linearly out from the center. The 3×3 groups of color images to the *left* show the individual components of the phase matrices Ψ_a, Ψ_b (*top two*) and phase matrix product, $\Psi_{\frac{a}{b}}$ (*bottom*). Here *green* means positive real, *red* means negative real, *blue* means positive imaginary and *yellow* means negative imaginary. The corresponding color coded phase images are shown to the *right*. The added *arrows* indicate the corresponding displacement estimate, $\mathbf{d}_\psi$, (*bottom right*). The *arrows* correctly show that the phase difference is pointing towards the center with magnitudes increasing linearly from the center. The outer *gray* circular area is where the difference in phase angle, θ, reaches π and the phase angle difference is interpreted as having the opposite sign

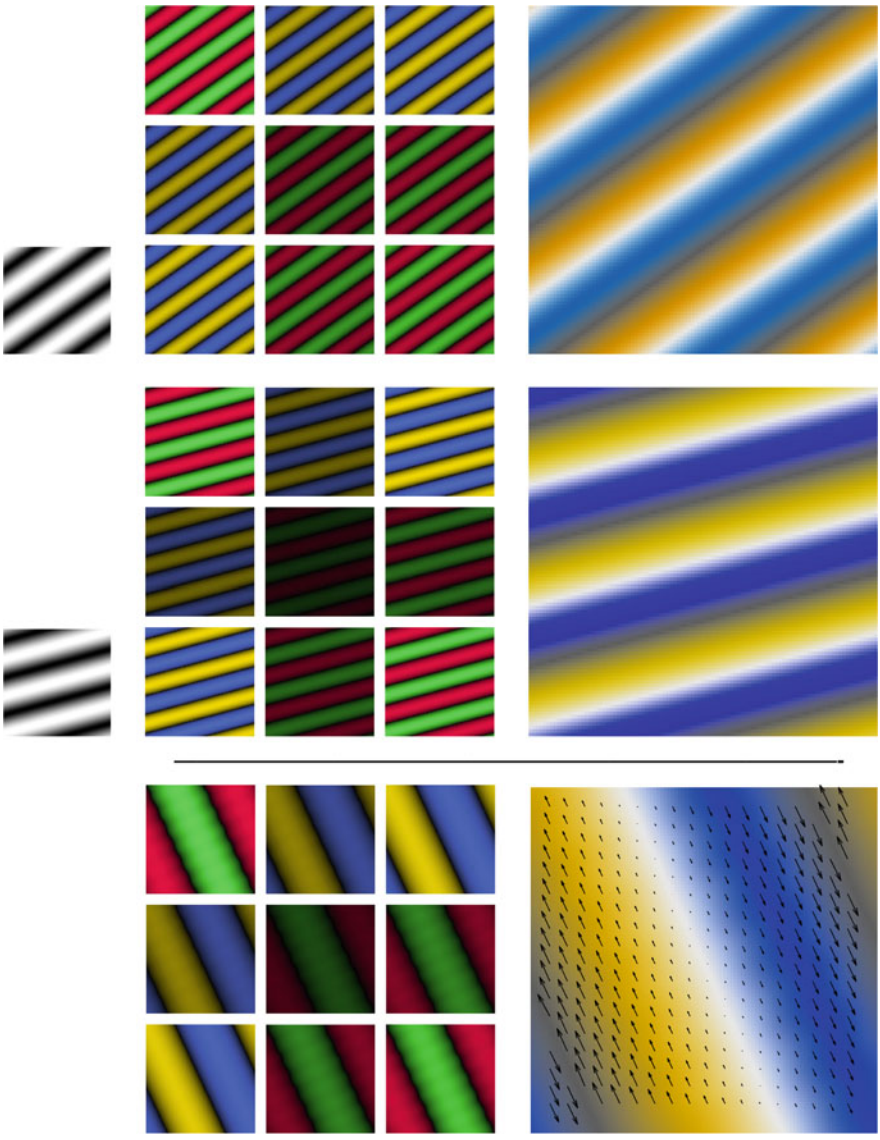


Fig. 14 Example using two sine waves with the same frequency but with different orientations, the orientation difference here is 20°. The layout is the same as in Fig. 13. The result shows that the estimated orientation of the motion corresponds to the average orientation present in the two input images, $\hat{\mathbf{s}}$. This is in accordance with Eq. (78) for small differences in orientation. Even though the orientation is relatively large the influence of the additional terms in Eq. (78) is hardly noticeable. The *white* linear structure going through the center of the result image (*bottom right*) indicates that the difference in the phase angle, θ is zero. The colors and the *arrows* correctly show that the phase difference grows linearly perpendicular to the *white* 'center-line'. The outer *gray linear areas* are where the difference in phase angle, θ, reaches π and the phase angle difference is interpreted as having the opposite sign

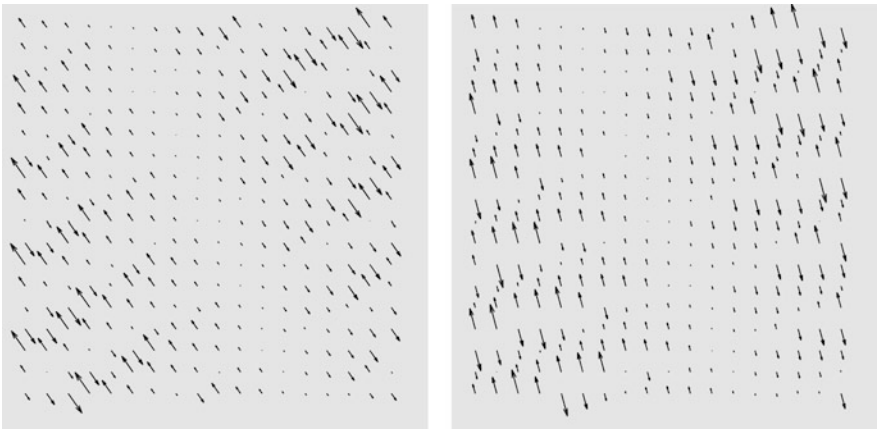


Fig. 15 Figure showing the result when applying the optical flow algorithm to the two *gray* images in Fig. 14. As the gradient in optical flow is taken from a single reference image there are two possible solutions, shown *left* and *right*. Even for these relatively simple cases the fundamental optical flow assumptions are violated and the result is very far from the correct displacement in most locations. The only situation that consistently gives the correct result is when the displacement is zero!

is highlighted. In motion estimation the motion field will, as a rule, have a much slower spatial variation then the local phase field. This situation is clearly shown in Figs. 13 and 14. Hence, a large low pass filter, that would remove important details of the phase field, can still be used without significant deterioration of the motion field. This will allow robust motion estimates to be attained in difficult and noisy situations.

The averaging operation can be expressed as a weighted summation where the coefficients, w_l sum to unity, i.e. $\sum w_l = 1$. Using a bar to denote a weighted average we can express the local average of the phase matrix product as:

$$\begin{aligned}
\overline{\boldsymbol{\Psi}}_{\substack{a \\ b}} &= \sum_l w_l \, \boldsymbol{\Psi}_{al} \, \boldsymbol{\Psi}_{bl}^T \\
&= \begin{pmatrix} \sum_l w_l \cos(\Delta\theta_l) & i \sum_l w_l \sin(\Delta\theta_l) \; \hat{\mathbf{a}}_l{}^T \\ i \sum_l w_l \sin(\Delta\theta_l) \; \hat{\mathbf{a}}_l & \sum_l w_l \cos(\Delta\theta_l) \; \hat{\mathbf{a}}_l \hat{\mathbf{a}}_l{}^T \end{pmatrix} \\
&= \begin{pmatrix} \overline{\mathsf{c}} & i \; \overline{\mathbf{c}}^T \\ i \; \overline{\mathbf{c}} & \overline{\mathbf{C}} \end{pmatrix}
\end{aligned} \tag{83}$$

In the general case the weights, w_l, of the averaging filter can be both spatially variant and signal dependent.

4.4.1 Equal Orientations

Letting all orientations be equal, i.e. $\hat{\mathbf{a}}_l = \hat{\mathbf{a}}\,;\ \forall l$ and calculating the weighted average, $\bar{z}$, of the complex numbers, z_l, corresponding to the difference phase angles $\Delta\theta_l$ in Eq. (83), i.e.

$$\bar{z} = \sum_l w_l\, z_l = \sum_l w_l\, e^{i\,\Delta\theta_l} \tag{84}$$

we find the weighted average phase matrix product as:

$$\begin{aligned} \overline{\boldsymbol{\Psi}}_{\substack{a \\ b}} &= \begin{pmatrix} \mathrm{Re}(\bar{z}) & i\ \mathrm{Im}(\bar{z})\ \hat{\mathbf{a}}^T \\ i\ \mathrm{Im}(\bar{z})\ \hat{\mathbf{a}} & \mathrm{Re}(\bar{z})\ \hat{\mathbf{a}}\,\hat{\mathbf{a}}^T \end{pmatrix} \\ &= r\ \boldsymbol{\Psi}(\Delta\theta,\ \hat{\mathbf{a}}) \end{aligned} \tag{85}$$

where $r \leq 1$ is a scalar attaining unity only if all z_l are equal. In fact the value of r will be directly related to the width of the angular distribution of z_l. Again we have a direct analog to complex numbers and it should be noted that, in general, the average phase angle (a bounded real number) will not be the same as the phase angle of the average phase matrix, i.e.

$$\Delta\theta \neq \overline{\Delta\theta_l} \tag{86}$$

4.4.2 General Averages

Analyzing the behavior of monomial phase matrix averages in the general case is beyond the scope of this chapter. Suffice to say that the averages will, due to the well-behaved representation, be continuous with respect to all entries and that the sub matrix $\overline{\mathbf{C}}$ will hold information about the directional distribution of $\hat{\mathbf{a}}_l$. For example, neighborhoods with a high degree of curvature will result in a $\overline{\mathbf{C}}$ having more than one significantly non zero eigenvalue.

5 Conclusion

We have presented a novel matrix representation of multidimensional phase that has a number of advantages. In contrast to previously suggested phase representations it is shown to be globally isometric, i.e. the metric of the representation is invariant to shifts along the phase matrix manifold. The proposed phase estimation approach uses spherically separable monomial filter of orders 0, 1 and 2 which naturally extends to N dimensions. For 2-dimensional simple signals the representation has the topology of a Klein bottle. Further, we have defined a symmetric conjugate

phase matrix product which retains the structure of the phase matrix representation. This product gives a phase difference matrix in a manner similar to the complex conjugate product of complex numbers. We have also shown that the phase matrix representation allows meaningful averages to be calculated as simple weighted summations. Some important advantages of the proposed representation has been demonstrated in two motion estimation examples.

We have to some extent investigated the mathematical properties of our new matrix phase representation. However, as pointed out by one of the reviewers, this matrix representation no doubt merits further investigation in particular regarding geometrical and group theoretical aspects.

Acknowledgements The authors would like to thank Mats Andersson, Michael Felsberg, Gustaf Johansson and Jens Sjölund for valuable discussions and proof reading, and Anders Brun for demonstrating the Klein bottle phase structure of oriented patches using his *LogMap* manifold learning algorithm. We also gratefully acknowledge the support from the Swedish Research Council grants 2011–5176, 2012–3682 and NIH grants R01MH074794, P41RR013218, and P41EB015902.

References

1. Riesz, M.: Sur les fonctions conjuge'es. Math. Zeit. **27**, 218–244 (1927)
2. Zernike, F.: Diffraction theory of the cut procedure and its improved form, the phase contrast method. Physica **1**, 689–704 (1934)
3. Bhatia, A.B., Wolf, E.: On the circle polynomials of Zernike and related orthogonal sets. Proc. Camb. Philos. Soc. **50**, 40–48 (1954)
4. Gabor, D.: Theory of communication. J. Inst. Electr. Eng. **93**(26), 429–457 (1946)
5. Stark, H.: An extension of the Hilbert transform product theorem. Proc. IEEE **59**(9), 1359–1360 (1971)
6. Bedrosian, E., Stark, H.: Comments on "an extension of the Hilbert transform product theorem". Proc. IEEE **60**(2), 228–229 (1972)
7. Bracewell, R.: The Fourier Transform and Its Applications, 2nd edn. McGraw-Hill, New York (1986)
8. Knutsson, H.: Filtering and reconstruction in image processing. PhD thesis, Linköping University, Sweden (1982). Diss. No. 88
9. Wilson, R., Knutsson, H.: A multiresolution stereopsis algorithm based on the Gabor representation. In: 3rd International Conference on Image Processing and Its Applications, Warwick, July 1989, pp. 19–22. IEE. ISBN:0 85296382 3, ISSN:0537-9989
10. Haglund, L., Knutsson, H., Granlund, G.H.: On phase representation of image information. In: The 6th Scandinavian Conference on Image Analysis, Oulu, June 1989, pp. 1082–1089
11. Haglund, L., Knutsson, H., Granlund, G.H.: Scale analysis using phase representation. In: The 6th Scandinavian Conference on Image Analysis, Oulu, June 1989, pp. 1118–1125
12. Fleet, D.J., Jepson, A.D.: Computation of component image velocity from local phase information. Int. J. Comput. Vis. **5**(1), 77–104 (1990)
13. Fleet, D.J., Jepson, A.D., Jenkin, M.R.M.: Phase-based disparity measurement. CVGIP Image Underst. **53**(2), 198–210 (1991)
14. Hahn, S.L.: Multidimensional complex signals with single-orthant spectra. Proc. IEEE **80**(8), 1287–1300 (1992)

15. Knutsson, H.: Representing local structure using tensors. In: The 6th Scandinavian Conference on Image Analysis, Oulu, pp. 244–251, June 1989. Report LiTH–ISY–I–1019, Computer Vision Laboratory, Linköping University, Sweden (1989)
16. Knutsson, H., Westin, C.-F., Granlund, G.H.: Local multiscale frequency and bandwidth estimation. In: Proceedings of the IEEE International Conference on Image Processing, Austin, Nov 1994, pp. 36–40. IEEE. (Cited in Science: Vol. 269, 29 Sept. 1995)
17. Granlund, G.H., Knutsson, H.: Signal Processing for Computer Vision. Kluwer, Dordrecht/Boston (1995). ISBN:0-7923-9530-1
18. Felsberg, M., Sommer, G.: The monogenic signal. IEEE Trans. Signal Process. **49**(12), 3136–3144 (2001)
19. Bulow, T., Sommer, G.: Hypercomplex signals-a novel extension of the analytic signal to the multidimensional case. IEEE Trans. Signal Process. **49**(11), 2844–2852 (2001)
20. Farnebäck, G.: Polynomial expansion for orientation and motion estimation. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden (2002). Dissertation No. 790, ISBN:91-7373-475-6
21. Felsberg, M.: Disparity from monogenic phase. In: DAGM Symposium Mustererkennung, Zurich. Volume 2449 of Lecture Notes in Computer Science, pp. 248–256. Springer, Heidelberg (2002)
22. Felsberg, M., Sommer, G.: Image features based on a new approach to 2D rotation invariant quadrature filters. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) Computer Vision – ECCV 2002, Copenhagen. Volume 2350 of Lecture Notes in Computer Science, pp. 369–383. Springer (2002)
23. Knutsson, H., Andersson, M.: Implications of invariance and uncertainty for local structure analysis filter sets. Signal Process. Image Commun. **20**(6), 569–581 (2005)
24. Felsberg, M., Duits, R., Florack, L.: The monogenic scale space on a rectangular domain and its features. Int. J. Comput. Vis. **64**(2–3), 187–201 (2005)
25. Felsberg, M., Jonsson, E.: Energy tensors: quadratic, phase invariant image operators. In: Pattern Recognition, Vienna. Volume 3663 of Lecture Notes in Computer Science, pp. 493–500 (2005)
26. Knutsson, H., Andersson, M.: Morphons: segmentation using elastic canvas and paint on priors. In: IEEE International Conference on Image Processing (ICIP'05), Genova, Sept 2005
27. San Jose Estepar, R.: Local structure tensor for multidimensional signal processing: applications to medical image analysis. PhD thesis, University of Valladolid, Spain (2005)
28. Wietzke, L., Sommer, G.: The conformal monogenic signal. In: Proceedings of the 30th DAGM Symposium on Pattern Recognition, Munich, pp. 527–536. Springer, Berlin/Heidelberg (2008)
29. Knutsson, H., Westin, C.-F., Andersson, M.: Structure tensor estimation – introducing monomial quadrature filter sets. In: Laidlaw, D., Vilanova, A. (eds.) New Developments in the Visualization and Processing of Tensor Fields. Springer, Dagstuhl (2012)
30. Fleischmann, O., Wietzke, L., Sommer, G.: Image analysis by conformal embedding. J. Math. Imaging Vis **40**(3), 305–325 (2011)
31. Duits, R., Fuhr, H., Janssen, B., Bruurmijn, M., Florack, L., van Assen, H.: Evolution equations on Gabor transforms and their applications. Appl. Comput. Harmon. Anal. Linköping Studies in Science and Technology, Dissertations (2012), pp. 48–55. ISSN 0345-7524; 1157, ISBN 978-91-85715-02-2
32. Brun, A.: Manifolds in image science and visualization. PhD thesis, The Institute of Technology, Medical Informatics, Linkoping University (2007)

Order Based Morphology for Color Images via Matrix Fields

Bernhard Burgeth and Andreas Kleefeld

Abstract Mathematical morphology is a successful branch of image processing with a history of more than four decades. Its fundamental operations are dilation and erosion, which are based on the notion of supremum and infimum with respect to an order. From dilation and erosion one can build readily other useful elementary morphological operators and filters, such as opening, closing, morphological top-hats, derivatives, and shock filters. Such operators are available for grey value images, and recently useful analogs of these processes for matrix-valued images have been introduced by taking advantage of the so-called Loewner order. There is a number of approaches to morphology for vector-valued images, that is, color images based on various orders, however, each with its merits and shortcomings. In this chapter we propose an approach to (elementary) morphology for color images that relies on the existing order based morphology for matrix fields of symmetric 2×2-matrices. An RGB-image is embedded into a field of those 2×2-matrices by exploiting the geometrical properties of the order cone associated with the Loewner order. To this end a modification of the HSL-color model and a relativistic addition of matrices is introduced. The experiments performed with various morphological elementary operators on synthetic and real images provide results promising enough to serve as a proof-of-concept.

B. Burgeth (✉)
Faculty of Mathematics and Computer Science, Saarland University, 66041 Saarbrücken, Germany
e-mail: burgeth@math.uni-sb.de

A. Kleefeld
Faculty of Mathematics, Natural Sciences and Computer Science, Brandenburg University of Technology Cottbus, 03046 Cottbus, Germany
e-mail: kleefeld@tu-cottbus.de

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_4,

1 Introduction

Beginning with the path-breaking work of Matheron and Serra [21, 22] in the late sixties mathematical morphology has provided us with an abundance of tools and techniques to process real valued-images for applications ranging from medical imaging to geological sciences [15,23–25]. Erosion and dilation are the fundamental operations of grey scale morphology relying on the notion of minimum and maximum of real numbers. Since minimum and maximum in turn depend on the presence of an order, it is no surprise that morphology for vector valued i.e. color images does not always provide satisfactory results.

There have been numerous approaches how to extend the mathematical morphology framework to color or vector-valued images. The main ingredients for such a framework are ranking schemes and the proper notion of extremal operators such as supremum and infimum. Due to the lack of reasonable complete lattice for vectorial data numerous suggestions for ranking schemes (based on various notions of distances, projections, and real-valued transforms) have been made, for a well structured, comprehensive, in-depth, and still up-to-date survey the reader is referred to [2] and the extensive list of literature cited therein. In [9] and [13] a more historic account is presented, while for a study of the background in order theory see [3] and [10].

Depending on the choices made one obtains morphological transforms with specific properties. However, none of these attempts seems to have been accepted unanimously in the image processing community.

Somewhat surprising the situation for symmetric matrix valued images is not as hopeless as it might seem at first glance.

Here we consider a (symmetric) matrix valued images or *matrix field* F as a mapping

$$F : \Omega \subset \mathbb{R}^d \longrightarrow \mathrm{Sym}(n)$$

from a image domain Ω in $\mathbb{R}^d$ into the set $\mathrm{Sym}(n)$ of real symmetric $n \times n$-matrices.

There have been successful attempts to extend the operations of mathematical morphology to images with values in the set of positive definite real symmetric 2×2- or 3×3-matrices [5–7], since these types of data make a natural appearance in medical imaging as the output of diffusion tensor weighted magnetic resonance imaging (DT-MRI).

The advantage of the matrix valued setting over the vector valued one is the presence of a prominent order for symmetric matrices, the so-called Loewner order, and the richer algebraic structure of symmetric matrices.

Hence, the goal of this chapter is to present an approach to morphological operators for color images by embedding a color image suitably into a matrix field. Hence the morphology already developed for matrix fields will give rise to morphology for color images.

For the coding of a color image as a matrix field we will make use of a variant of the HSL-color space and the Loewner order cone for real symmetric 2×2-matrices. This novel concept can be applied to grey value images as well and indeed includes scalar (flat) morphology.

The structure of the article is as follows: In order to keep the chapter as self-contained as possible we devote the next section to a brief review of the grey scale morphological operations we aim to extend to the matrix-valued setting, starting from the basic erosion/dilation and reaching to the morphological equivalents of gradient, and Laplacian, and its use for shock filtering. In Sect. 3 we present the maximum and minimum operations for matrix-valued data and especially a three-dimensional representation of the Loewner order cone for 2×2- matrices. Section 4 contains the aforementioned embedding and an operation for symmetric matrices gleaned from the relativistic addition of velocities. We report the results of our experiments with various morphological operators applied to synthetic and real color images in Sect. 5. Section 6 offers concluding remarks and a short hint at future research.

2 A Glance at Scalar Morphology

In grey scale morphology a scalar function f represents an image: $f(x, y)$ with $(x, y) \in \mathbb{R}^2$. In this paper we restrict ourselves to flat grey scale morphology where a binary type of the so-called *structuring element* is used. It is nothing but a set B in $\mathbb{R}^2$ determining the neighborhood relation of pixels. Then grey scale *dilation* $\oplus$, resp., *erosion* $\ominus$ replaces the grey value of the image $f(x, y)$ by its supremum, resp., infimum within the mask B:

$$\begin{aligned}(f \oplus B)(x, y) &:= \sup\{f(x-x',\ y-y') \mid (x', y') \in B\}\ , \\ (f \ominus B)(x, y) &:= \inf\{f(x+x',\ y+y') \mid (x', y') \in B\}\ .\end{aligned}$$

By concatenation other operators are constructed such as *opening* and *closing*,

$$f \circ B := (f \ominus B) \oplus B\ , \qquad f \bullet B := (f \oplus B) \ominus B\ ,$$

the *white top-hat* and its dual, the *black top-hat*

$$\mathrm{WTH}(f) := f - (f \circ B)\ , \qquad \mathrm{BTH}(f) := (f \bullet B) - f\ ,$$

finally, the *self-dual top-hat*,

$$\mathrm{SDTH}(f) := (f \bullet B) - (f \circ B)\ .$$

The boundaries or edges of objects in an image are the loci of high grey value variations and those can be detected by gradient operators. Erosion and dilation are also the elementary building blocks of the basic morphological gradients, namely: The so-called *Beucher gradient*

$$\rho_B(f) := (f \oplus B) - (f \ominus B) \, .$$

It is an analog to the norm of the gradient $\|\nabla f\|$ if an image is considered as a differentiable function. Other useful approximations to $\|\nabla f\|$ are the *internal* and *external gradient*,

$$\rho_B^-(f) := f - (f \ominus B) \, , \qquad \rho_B^+(f) := (f \oplus B) - f \, .$$

A *morphological Laplacian* has been introduced in [26] as the morphological equivalent for the Laplace operator $\Delta F = \partial_{xx} F + \partial_{yy} F$ in the matrix valued setting. Following [8] we consider a variant given by the difference between external and internal gradient

$$\Delta_m F := \rho_B^+(F) - \rho_B^-(F) = (F \oplus B) - 2 \cdot F + (F \ominus B) \, ,$$

thus representing the second derivative $\partial_{\eta\eta} f$ where η denotes the direction of the steepest slope.

Since $\Delta_m f$ is matrix-valued, $\mathrm{trace}(\Delta_m f)$ will provide us with useful information: Regions where $\mathrm{trace}(\Delta_m F) \leq 0$ can be viewed as the influence zones of maxima while those areas with $\mathrm{trace}(\Delta_m F) \geq 0$ are influence zones of minima. Hence it allows us to distinguish between influence zones of minima and maxima in the image F. This is decisive for the construction of so-called *shock filters*.

The basic idea underlying shock filtering is applying either a dilation or an erosion to an image, depending on whether the pixel is located within the influence zone of a minimum or a maximum [19]:

$$\delta_B(f) := \begin{cases} f \oplus B & \text{if} \quad \mathrm{trace}(\Delta_m f) \leq 0 \, , \\ f \ominus B & \text{else} \, . \end{cases} \tag{1}$$

The shock filter expands local minima and maxima at the cost of regions with intermediate grey values. When iterated, experimental results in grey scale morphology suggest that a non-trivial steady state exists characterized by a piecewise constant segmentation of the image.

In the scalar case the zero-crossings $\Delta f = 0$ can be interpreted as edge locations [14,17,20]. We will also use the trace of the morphological Laplacian in this manner to derive an edge map.

However, if we consider a matrix field F, $\Delta_m F$ is matrix-valued, but the scalar value of $\mathrm{trace}(\Delta_m F)$ provides us with the information necessary for the switch criterion in the matrix valued shock filter.

3 Loewner Order: Maximal and Minimal Matrices

Morphology rests on the fundamental notions of supremum and infimum with respect to an order. This is true for matrix fields as well.

The so-called *Loewner order* is a natural partial order on $\mathrm{Sym}(n)$, defined via the cone of positive semidefinite matrices $\mathrm{Sym}^+(n)$ by

$$A, B \in \mathrm{Sym}(n): \quad A \geq B :\Leftrightarrow A - B \in \mathrm{Sym}^+(n),$$

i.e. if and only if $A - B$ is positive semidefinite.

This partial order is *not* a lattice order, that is, there is no notion of a unique supremum and infimum with respect to this order [4]. Nevertheless, given any finite set of symmetric matrices $\mathscr{A} = \{A_1, \ldots, A_n\}$, we will be able to identify suitable maximal, resp., minimal matrices

$$\overline{A} := \sup \mathscr{A} \quad \text{resp.,} \quad \underline{A} := \inf \mathscr{A} .$$

Since we will consider images with three color components we may restrict ourselves from now on to the case of 2×2-matrices in $\mathrm{Sym}(2)$ which offer already three degrees of freedom. The procedure to find these extremal matrices for a set $\mathscr{A}$ is as follows: The cone $\mathrm{Sym}^+(2)$ can be represented in 3D using the bijection

$$A := \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \longleftrightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 2\beta \\ \gamma - \alpha \\ \gamma + \alpha \end{pmatrix}, \text{ resp., } \frac{1}{\sqrt{2}} \begin{pmatrix} z - y & x \\ x & z + y \end{pmatrix} \longleftrightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix} . \quad (2)$$

This mapping creates an isometrically isomorphic image of the cone $\mathrm{Sym}^+(2)$ in the Euclidean space $\mathbb{R}^3$ given by $\{(x, y, z)^\top \in \mathbb{R}^3 | \sqrt{x^2 + y^2} \leq z\}$ and is depicted in Fig. 1a. For $A \in \mathrm{Sym}(2)$ the set $P(A) = \{Z \in \mathrm{Sym}(2) | A \geq Z\}$ denotes the penumbral cone or penumbra for short of the matrix A. It corresponds to a cone with vertex in A and a circular base in the $x - y$-plane:

$$P(A) \cap \{z = 0\} = \text{ circle with center } (\sqrt{2}\beta, \frac{\gamma - \alpha}{\sqrt{2}}) \text{ and radius } \frac{\mathrm{trace}(A)}{\sqrt{2}} .$$

Considering the associated penumbras of the matrices in $\mathscr{A}$ the search for the *maximal matrix* $\overline{A}$ amounts to determine the smallest penumbral cone covering all the penumbras of $\mathscr{A}$ tightly, see Fig. 1b. One realises that the height of a penumbra measured from the $x - y$-plane is equal to the radius of its base, namely $\frac{\mathrm{trace}(A)}{\sqrt{2}}$. Hence a penumbra is already uniquely determined by the circle constituting its base. This implies that the search for a maximal matrix comes down to finding the smallest circle enclosing the base-circles of the matrices in $\mathscr{A}$. This is a non-trivial problem in computer graphics. A numerical solution for finding the smallest circle enclosing

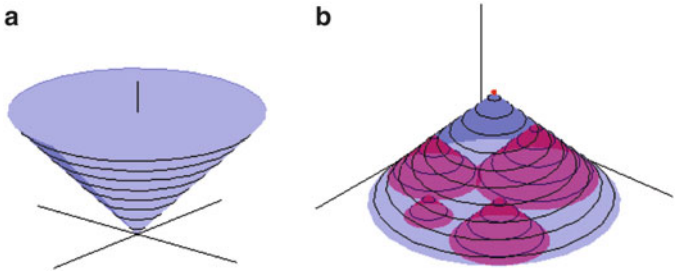


Fig. 1 (**a**) Image of the Loewner cone $\mathrm{Sym}^+(2)$. (**b**) Cone covering four penumbras of other matrices. The tip of each cone represents a symmetric 2×2 – matrix in $\mathbb{R}^3$. For each cone the radius and the height are equal

the sampled basis circles has been implemented in C++ by Gärtner [11] and was used in [7] and [6]. However, in our case we employ the implementation of an efficient subgradient method detailed in [27] for the calculation of the smallest circle enclosing them. This gives us the smallest covering cone and hence the maximal matrix $\overline{A}$. For technical reasons in our later application we will not apply the above reasoning directly to the matrices $A_1, \ldots, A_n$ but to their shifted versions $\mathscr{A}+I := A_1+I, \ldots, A_n+I$ with the unit matrix I taking advantage of the relation

$$\overline{A} = \sup(A_1, \ldots, A_n) = \sup(A_1 + I, \ldots, A_n + I) - I = \overline{A + I} - I \ .$$

A suitable *minimal matrix* $\underline{A}$ is obtained by means of the formula

$$\underline{A} = I - (\sup(I - A_1, \ldots, I - A_n))$$

inspired by the relation $\min(a_1, \ldots, a_n) = 1 - (\max(1 - a_1, \ldots, 1 - a_n))$ valid for real numbers $a_1, \ldots, a_n$. For $i = 1, \ldots, n$ we have $\underline{A} \leq A_i \leq \overline{A}$ with respect to the Loewner order. We emphasise that $\overline{A}$ and $\underline{A}$ depend continuously on $A_1, \ldots, A_n$ by their construction. Also the rotational invariance is preserved, since the Loewner order is already rotational invariant: $A \geq B \iff UAU^\top \geq UBU^\top$ holds for any orthogonal matrix U. Nevertheless, the definitions of the matrices $\overline{A}$ and $\underline{A}$ are still meaningful for matrices that are not positive definite as long as they have a non-negative trace (since it corresponds to a radius in the construction above). It also becomes evident from their construction that in general neither $\overline{A}$ nor $\underline{A}$ coincide with any of the A_i.

With these essential notions of suitable maximal and minimal matrices $\overline{A}$ and $\underline{A}$ at our disposal the definitions of the higher morphological operators carry over essentially verbatim, with one exception:
The morphological Laplacian Δ_m as defined in Sect. 2 is a matrix. In Eq. (1) we used the *trace* of the morphological Laplacian to steer the interwoven dilation-erosion process, and to create an edge map.

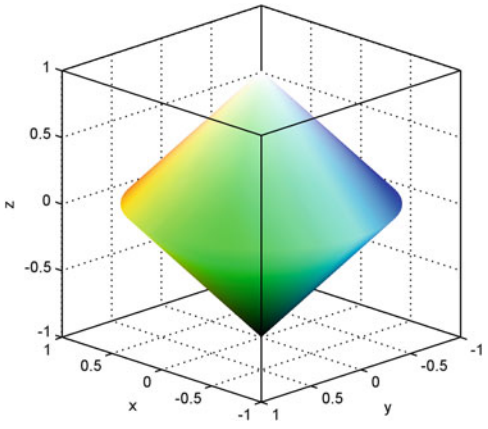


Fig. 2 HCL̃-bicone: Bicone for the HCL̃ color model

4 Color Images as Matrix Fields

One of the most common models is the RGB color space where each color appears in its primary components red, green, and blue. After standard normalizations the color space is represented by the unit cube in a Cartesian coordinate system, see [12]. Deeper inside in the use of color in sciences is provided in [18]. Closer to the human perception process is the HSL (or HSI) color model describing a color object by its hue, saturation and brightness resp. luminance. It is a popular cylindrical-coordinate representation of points in an RGB color model (see [1, Algorithm 8.6.3] for the conversion). If one replaces in this model the coordinate saturation by the so-called *chroma* one arrives at a modified version of the HSL-model which we call HCL̃ color model. Its representation is given by a bicone, depicted in Fig. 2.

To be more specific:

$$\tilde{L} = 2L - 1 \,,$$

and the chroma is obtained via

$$C = \max\{R, G, B\} - \min\{R, G, B\} \,.$$

Hence, any point (x, y, z) of the bicone can be obtained via

$$\begin{aligned} x &= C \cdot \cos(2\pi \cdot H) \,, \\ y &= C \cdot \sin(2\pi \cdot H) \,, \\ z &= \tilde{L} \,. \end{aligned}$$

Note that the ranges of all components lie in the unit interval. This provides us with an one-to-one transformation of the HCL̃ color space to the RGB color space. It is

now apparent, that this HC$\tilde{\text{L}}$-bicone corresponds via (2) directly to the order interval

$$[-I, I]_L := \{A \in \text{Sym}(2) \mid -I \leq A \leq I\} .$$

Hence each matrix in $[-I, I]_L$ corresponds uniquely to a point in the bicone and each point in the bicone represents uniquely a color. In total this establishes the desired continuous one-to-one correspondence of the matrices in $[-I, I]_L$ with the colors in the HC$\tilde{\text{L}}$ (and from there to the standard RGB space, if so desired),

$$\Psi : \text{HC}\tilde{\text{L}} \subset IR^3 \longrightarrow [-I, I]_L \subset \text{Sym}(2) .$$

Exploiting this correspondence one obtains, for example, the supremum of two colors $c_1, c_2 \in \text{HC}\tilde{\text{L}}$ by transforming them into the matrices $\Psi(c_1), \Psi(c_2) \in [-I, I]_L$, then taking the supremum $\sup(\Psi(c_1), \Psi(c_2))$ of these two matrices which is then transformed back to the new 'supremum color'

$$\sup(c_1, c_2) := \Psi^{-1}\left(\sup(\Psi(c_1), \Psi(c_2))\right) .$$

Ψ^{-1} combines the mapping (2), namely

$$\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \to \frac{1}{\sqrt{2}} \begin{pmatrix} 2\beta \\ \gamma - \alpha \\ \gamma + \alpha \end{pmatrix} =: \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

with a transform into polar coordinates via

$$\begin{aligned} H &= \frac{1}{2\pi} \text{arg(y, x)} , \\ C &= \sqrt{x^2 + y^2} , \\ \tilde{L} &= z , \end{aligned}$$

with a principal value of an appropriate argument function. The luminance L is obtained via $L = (\tilde{L} + 1)/2$ while the saturation is given by $S = 0$ if $C = 0$, otherwise $S = C/(1 - |2L - 1|)$.

Thus, having obtained those HSL-values, we now convert them to the normalized RGB-values (see [1, Algorithm 8.6.4] for the conversion).

The infimum of two colors is treated analogously. Hence by applying the same rationale not only the fundamental processes of dilation and erosion can be transferred from matrix fields to color images but other basic morphological operations as well.

However, two major difficulties arise:

1. Due to the non-polygonal nature of the Loewner order cone it can happen, that the supremum (and likewise the infimum) of matrices is outside the interval $[-I, I]_L$, that is, the corresponding HCL- and RGB-value do not exist.

2. The design of the morphological top-hat or derivative operators requires taking the differences/sums of matrices in $[-I, I]_L$. Taking this differences/sums in the standard way of matrix algebra results in matrices outside $[-I, I]_L$, again entailing non-existing corresponding HCL- and RGB-values.

We overcome the first difficulty by a rescaling of the corresponding matrices: Instead of the matrix $\Psi(c)$ associated to a color $c \in$ HCL we consider the matrix $\frac{\sqrt{2}}{2}\Psi(c)$ involving the scaling factor $d = \sqrt{2}/2$.

To resolve the second problem we extend Einstein's rule for the addition velocities in the theory of Special Relativity to symmetric matrices:

$$A +\!\!+ B := (A + B) \circ \left(1 + \frac{1}{2} A \circ B\right)^{-1} \tag{3}$$

where the symmetric matrix product '$\circ$' defined by

$$A \circ B := \frac{1}{2}(AB + BA)$$

is the Jordan product of A and B. This type of addition is inspired from the fact that the interval of real numbers $[-d, d] \subset \mathbb{R}$ with $d = \frac{\sqrt{2}}{2}$ equipped with the relativistic addition

$$a +\!\!+ b := \frac{a + b}{1 + \frac{a \cdot b}{d^2}}$$

of numbers $a, b \in [-d, d]$ establishes a commutative group [16], hence allows naturally for subtraction as well. This relativistic rule of addition/subtraction of matrices will lead to matrices often located in the lower part of the bicone causing the resulting color to tend towards grey/black. This effect is responsible for the very dominant tone of grey in images processed with operators invoking a subtraction, such as morphological top-hats and derivatives. Subsequent experimental results confirm this reasoning.

5 Experimental Results

In this section, the morphological operators designed above will be applied to synthetic as well as natural color images of various sizes.

In all the experiments we use a cross-shaped structuring element consisting of five pixels centered at the middle pixel. Each of the images is extended by one layer of mirrored boundary values.

In the first experiments we confirm that our color-morphological operators applied to image in Fig. 3a in principle act as regular morphological operators on

Fig. 3 Original 8 × 8-image test images. (**a**) Image 1. (**b**) Image 2. (**c**) Image 3

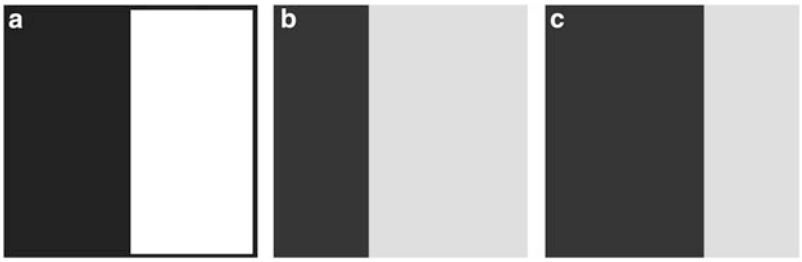


Fig. 4 Dilation and erosion applied to bipartite *black-and-white image*. (**a**) Original 8 × 8-image. (**b**) Dilation. (**c**) Erosion

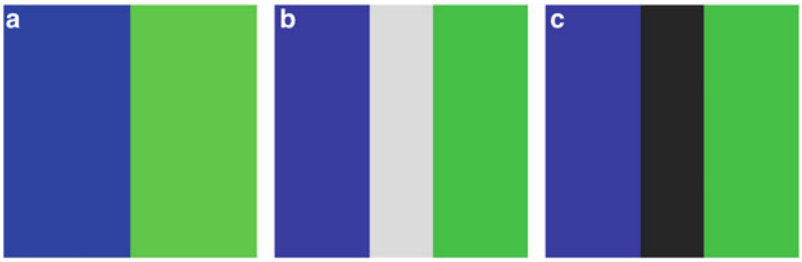


Fig. 5 Dilation and erosion applied to bipartite *blue-and-green image*. (**a**) Original 8 × 8-image. (**b**) Dilation. (**c**) Erosion

black-and-white images. As expected dilation and erosion result in an accurate shift of the inner object front, see Fig. 4a. However, the scaling causes a slight shift of black and white towards grey: The black part which is represented by the RGB-values $[0, 0, 0]$ becomes dark grey $[37, 37, 37]$. The white part which is represented by the RGB-values $[255, 255, 255]$ turns into light grey $[218, 218, 218]$.

An image of the same size 8 × 8, but with a blue-colored (RGB = $[0, 0, 255]$) left and a green-colored (RGB = $[0, 255, 0]$) right side undergoes both an dilation and an erosion. Both colors are located in the $x - y$-plane relatively far apart on the boundary of the HC$\tilde{\text{L}}$ bicone, hence the maximal and minimal matrices are representing almost white and black respectively. This accounts for the light grey

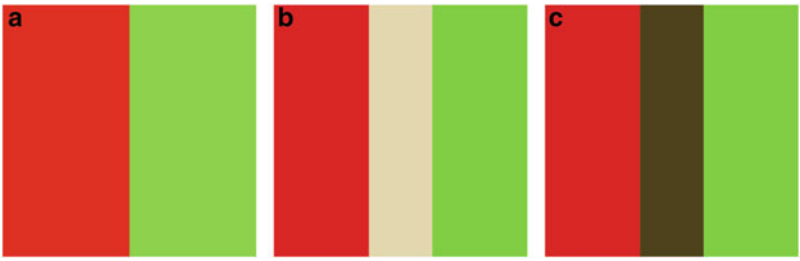


Fig. 6 Bipartite color image for which the rescaling of the bicone is necessary. (**a**) Original 8×8-image. (**b**) Dilation. (**c**) Erosion

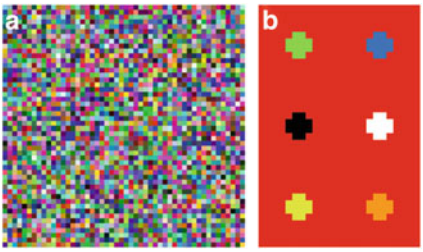


Fig. 7 Original 50×50- and 36×24-test images. (**a**) Image 4. (**b**) Image 5

resp. dark grey center section in the dilated resp. eroded images in Figs. 5b, c. Note that due to the scaling the colors are slightly faded. We repeat the experiment with a bipartite image whose halves are colored with RGB-values $[255, 0, 0]$ ($\mathrm{HC\tilde{L}} = [0, 1, 0]$) and $[128, 255, 0]$ ($\mathrm{HC\tilde{L}} = [1/4, 1, 0]$). Their representing $\mathrm{HC\tilde{L}}$-values are located at the boundary of the base of the bicone in the $x - y$-plane at a 90° angle. This represents a worst-case-scenario in the sense that the maximum of the two color-representing cones does not lie completely within the bicone, unless one uses the above mentioned scaling prior to taking the maximum. The same is true for taking the minimum. In total this results in a grey tone of the appearing colors, see Figs. 6b, c.

Now, we apply dilation and erosion as well as their concatenations opening and closing to a 50×50-image with random RGB-coloring of its pixels, see Figs. 7a and 8a respectively. As expected from the color distribution within the bicone erosion Fig. 8c entails a darkening of the image while dilation Fig. 8b accounts for its overall brightening. Inspecting Fig. 8d, e it becomes apparent that opening and closing lead to a coarsening of the image making the structuring element a clearly discernible shape.

Let us consider an RGB-image of resolution 36×24 containing six circles with the RGB colors $[128, 255, 0]$, $[0, 128, 255]$, $[0, 0, 0]$, $[255, 255, 255]$, $[235, 249, 18]$ and $[249, 155, 18]$ above a red background ($[255, 0, 0]$). In Figs. 7b and 9a respectively we show the constructed image, whereas in Fig. 9b, c we see the results for

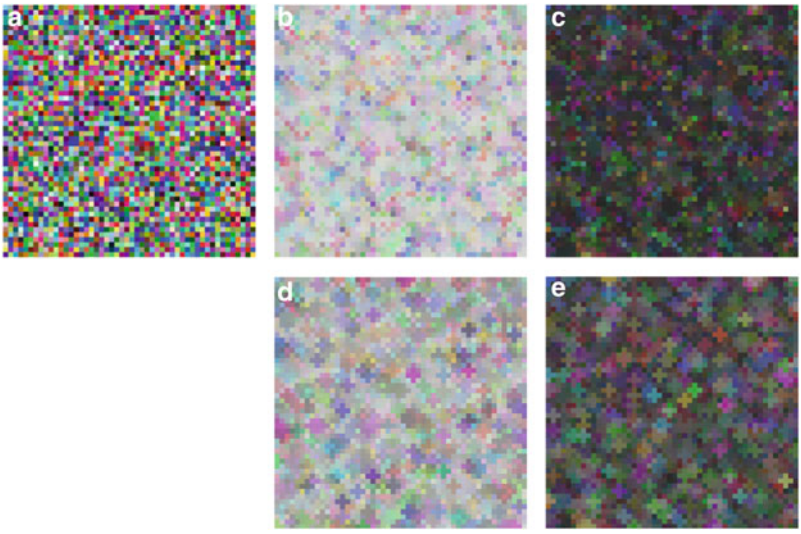


Fig. 8 Dilation, erosion, opening, and closing of an image with randomly colored pixels. (**a**) Original image of resolution 50 × 50. (**b**) Dilation. (**c**) Erosion. (**d**) Closing. (**e**) Opening

dilation and erosion. Comparing the results of opening and closing displayed in Fig. 9d, e with the original image we note its almost perfect recovery. Only the colors lost a bit of their brilliance, again an effect of the aforementioned scaling.

The outcome of the operations of dilation, erosion, opening, and closing when applied to a natural image Figs. 10b and 11a respectively is depicted in Fig. 11b–e. The closing operation behaves indeed as a filter eliminating the slight pepper-noise in the original image. Both closing and opening operations preserve to some extent the texture-like structure of the forest in the picture while simultaneously diminishing the depicted clouds.

Another example of the effect of these for basic operations on natural images, see Fig. 10, is displayed in Fig. 12. Dilation causes the white stripes in the flag to widen and creates a slightly bright rim around the depicted building whereas erosion induces a widening of the red stripes and a greyish rim around the building.

When polluted with 5 % salt-noise, Fig. 12d, an opening removes it, Fig. 12f, as expected. The same is true with 5 % pepper-noise, Fig. 12e, eliminated by a closing operation, see Fig. 12g.

In the next experiment we investigate the effect of repeated dilation and erosion on both a synthetic and a natural image. In the synthetic image small structures assume more and more the shape of the structuring element, as expected. However, repeated dilations produce bright-colored rims at the cost of the objects itself, the black structure in Fig. 13a even vanishes and blends into the red background. It is the opposite with erosion: the eroded structures are first surrounded and then swallowed in a thickening dark rim and white structures disappear into the red background.

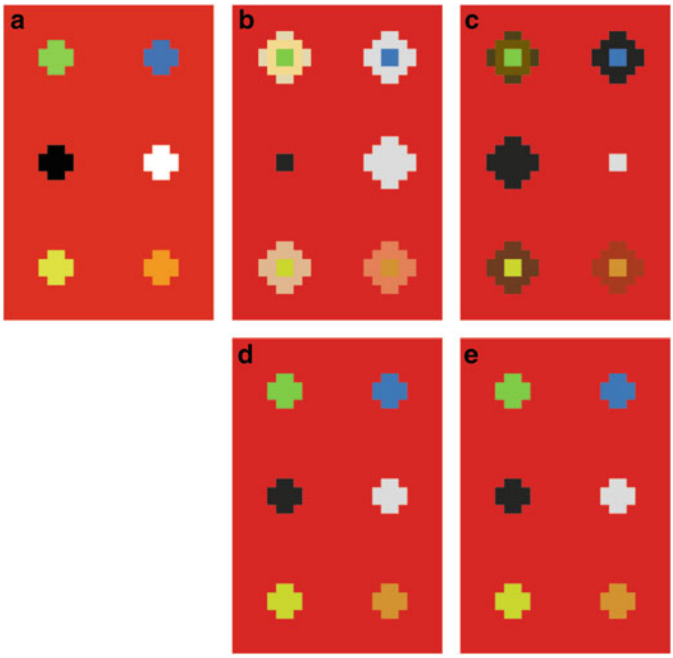


Fig. 9 Dilation, erosion, opening, and closing. (**a**) Original. (**b**) Dilation. (**c**) Erosion. (**d**) Closing. (**e**) Opening

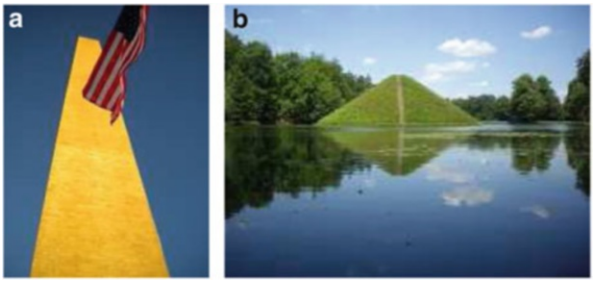


Fig. 10 Original 171 × 228- and 228 × 171-test images. (**a**) Image 6. (**b**) Image 7

In principle the same is true for natural images, prominent objects are getting surrounded by a thickening rim of light grey resp. dark grey tone when the number of dilations resp. erosions performed increases, as can be seen in Fig. 14. In the next set of experiments we turn our attention to the morphological derivatives: internal, external, and Beucher gradient, and the morphological Laplacian. They invoke a difference operation which is realized by means of the relativistic addition (3). The predominant tone in the derived image can be expected to be grey, the shade of color in the center of the bicone. Indeed, the difference operation underlying all

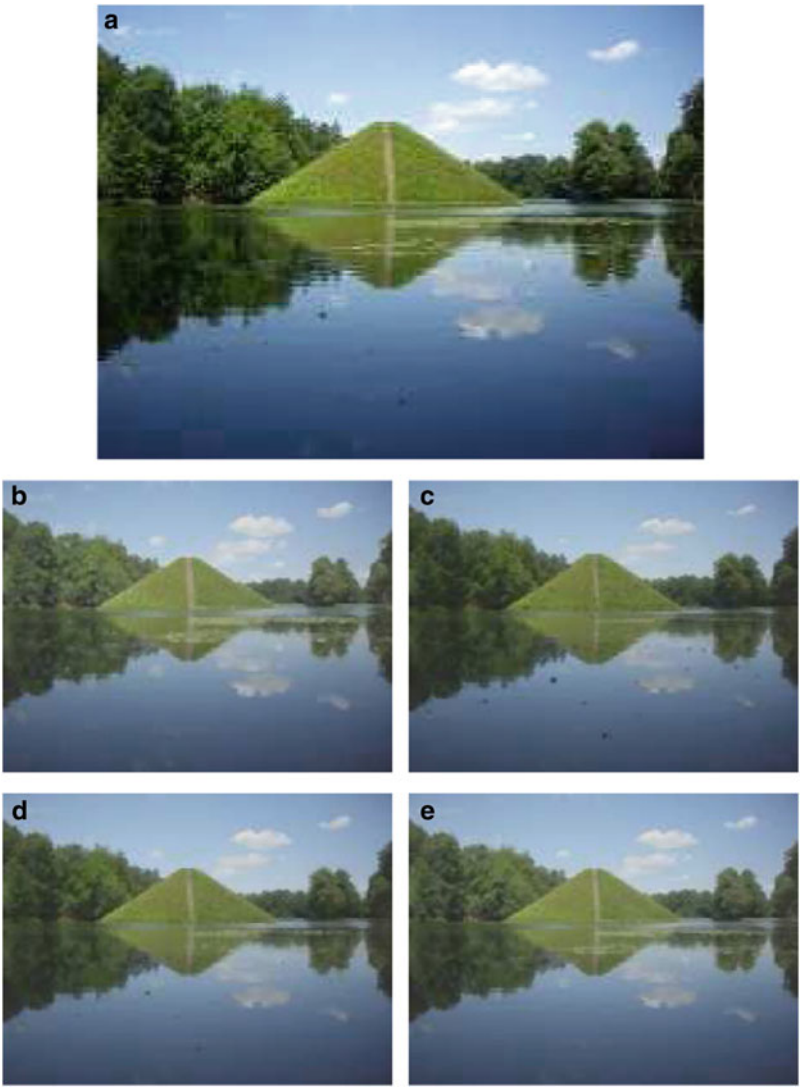


Fig. 11 Dilation, erosion, opening, and closing of a natural image. (**a**) Original 228 × 171-image. (**b**) Dilation. (**c**) Erosion. (**d**) Opening. (**e**) Closing

the derivatives leads to matrices close to the said center. Only where a considerable change by erosion and/or dilation occurs the subtraction entails a matrix somewhat outside the bicone center, that is a color close to the primary colors. The experiments with the matrix-based derivative operations corroborate this reasoning.

Nevertheless, it becomes apparent that the morphological gradients may serve as edge detectors, the Beucher gradient being slightly more dissipative than the other two one-sided ones (Figs. 15 and 16).

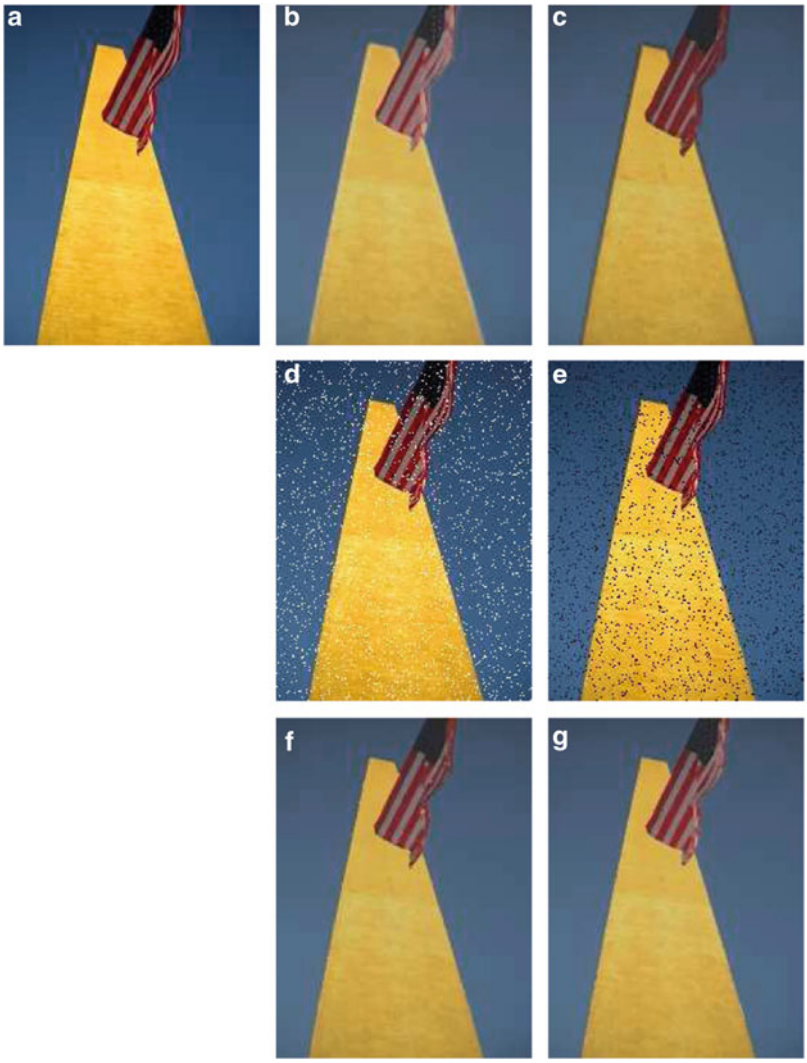


Fig. 12 Natural image. *Top*: Dilation, erosion. *Middle*: Image polluted with 5 %-salt noise and 5 %-pepper noise. *Bottom*: Removal of these noise types by opening and closing. (**a**) Original 171 × 228-image. (**b**) Dilation. (**c**) Erosion. (**d**) 5 % Salt noise added. (**e**) 5 % Pepper noise added. (**f**) Opening. (**g**) Closing

An obvious approach to morphology to color images is the channel-wise application of the morphological operations. That means, every RGB-image is decomposed into three scalar images and the morphological operation is applied to each of the three images separately. The resulting images are then re-combined to a now transformed RGB-image.

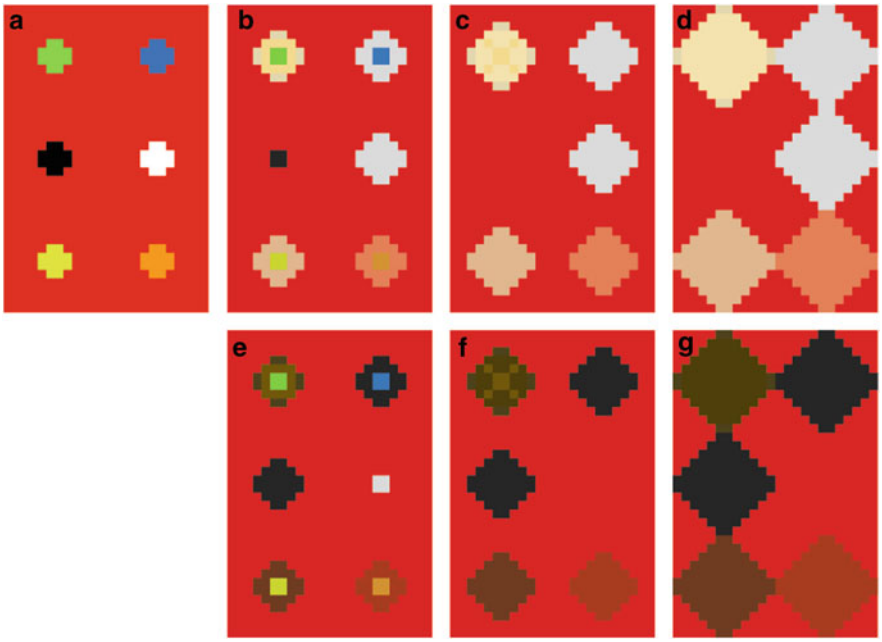


Fig. 13 Iterated dilations and iterated erosions. (**a**) Original. (**b**) Dilation 1×. (**c**) Dilation 2×. (**d**) Dilation 4×. (**e**) Erosion 1×. (**f**) Erosion 2×. (**g**) Erosion 4×

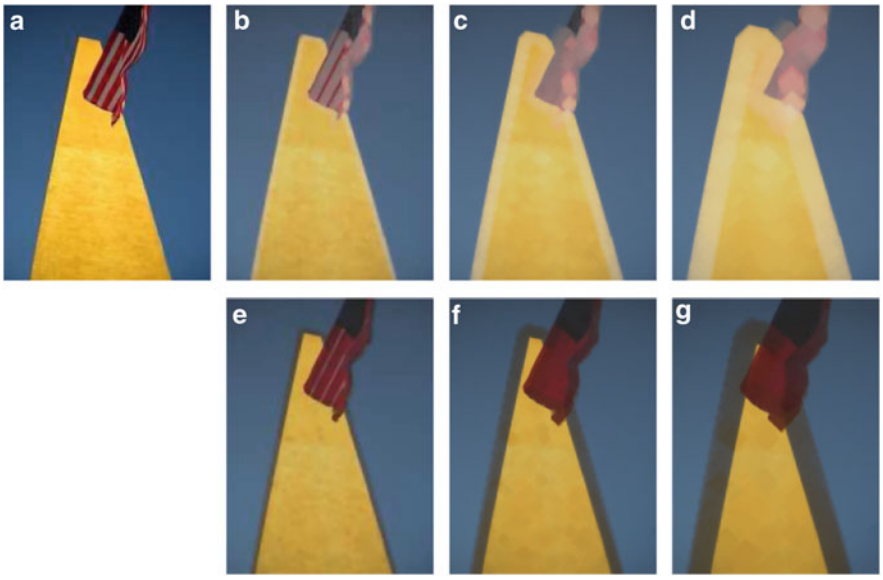


Fig. 14 Iterated dilations and iterated erosions based on the matrix field approach. (**a**) Original. (**b**) Dilation 2×. (**c**) Dilation 5×. (**d**) Dilation 10×. (**e**) Erosion 2×. (**f**) Erosion 5×. (**g**) Erosion 10×

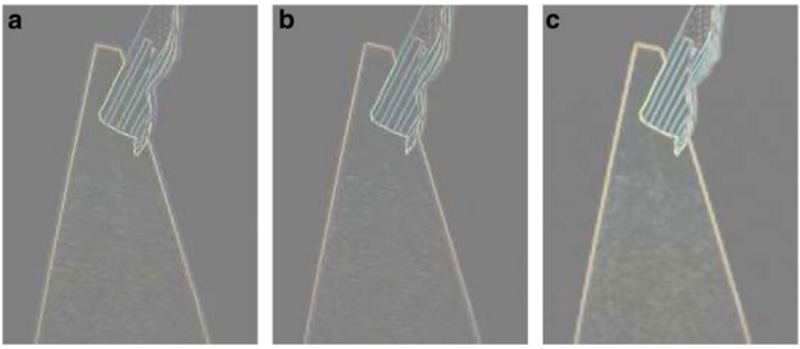


Fig. 15 Internal, external, and Beucher gradient applied to a natural image. (**a**) Internal. (**b**) External. (**c**) Beucher gradient

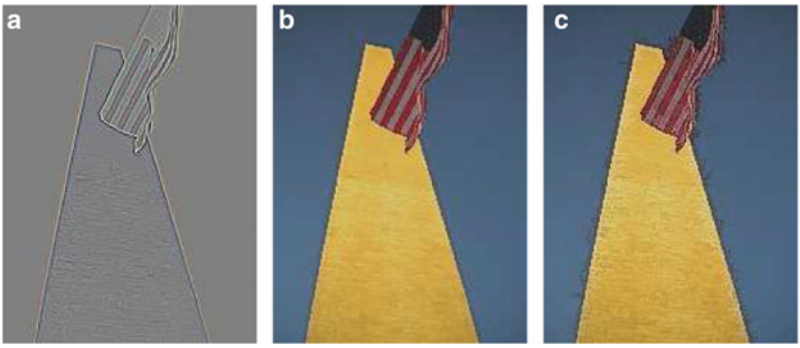


Fig. 16 The morphological Laplacian which provides the switch criterion for the shock filter applied once and 10 times to a natural image. (**a**) Morph. Laplacian. (**b**) Shockfilter 1×. (**c**) Shockfilter 10×

This works reasonably well as far as dilation, erosion, opening and closing with a small structuring element are concerned. The rim that is expected to get thicker with each iteration assumes a green color using the component-wise approach, see Fig. 17 which is different compared to the new approach, see Fig. 14.

Furthermore, these color disturbances become even more severe if a difference operation is involved, as the next examples exhibit, see Fig. 18. Trying a component-wise approach with the channels of the HC$\tilde{L}$-model is as fruitless as with the channels of the RGB-model. The other color phenomenon manifests itself clearly in Fig. 19, where we depicted only the HC$\tilde{L}$-channel-wise Beucher gradient and Laplacian.

So far we have refrained from presenting results of the application of the matrix-based morphological operations referred to as black top-hat, white top-hat and self-dual top-hat to color images. In fact, it is difficult to say what one should expect

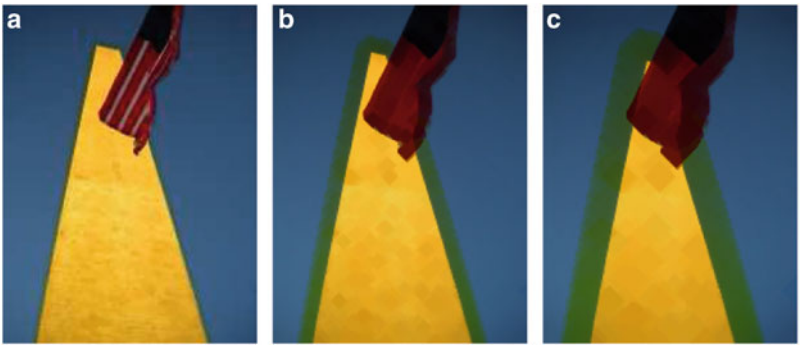


Fig. 17 Component-wise approach: Iteration of the erosion operation applied separately to each of the three channels of an RGB-image. The three resulting scalar images are then re-combined to the RGB-image. (**a**) Component-wise erosion 1×. (**b**) Component-wise erosion 5×. (**c**) Component-wise erosion 10×

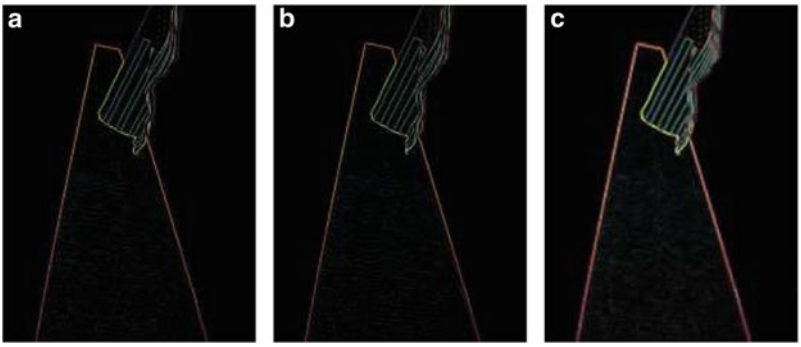


Fig. 18 Component-wise approach: The morphological operations of internal, external, and Beucher gradient applied to each of the three channels of an RGB-image separately followed by a recombination to a single RGB-image. (**a**) Component-wise internal gradient. (**b**) Component-wise external gradient. (**c**) Component-wise Beucher gradient

from a filter that is supposed to extract small black details from a (grey scale) image if it is applied to a color image.

In the experiments the predominant tone in the transformed image is again grey, the shade of color in the center of the bicone. The explanation for its appearance is practically the same as for the derivative operations since the top-hats are based on the difference operation as well: Only where a considerable change by opening and/or closing occurs the subtraction leads to matrices out of the bicone center, meaning, to colors close to primary colors. This is confirmed by the experiments with the matrix-based top-hat operations, see Fig. 20. However, the results of the component-wise RGB-based approach are not satisfactory either as can be seen in Fig. 21.

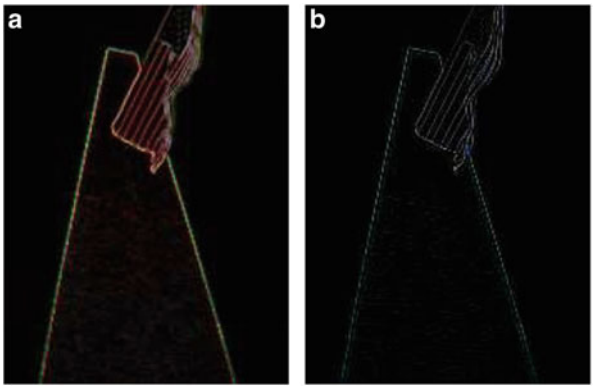


Fig. 19 HC$\tilde{\text{L}}$-component-wise approach: The morphological operations Beucher gradient and Laplacian applied to each of the three channels of an HC$\tilde{\text{L}}$-image separately followed by a recombination to a single HC$\tilde{\text{L}}$-image. (**a**) HC$\tilde{\text{L}}$-component-wise Beucher gradient. (**b**) HC$\tilde{\text{L}}$-component-wise Laplacian

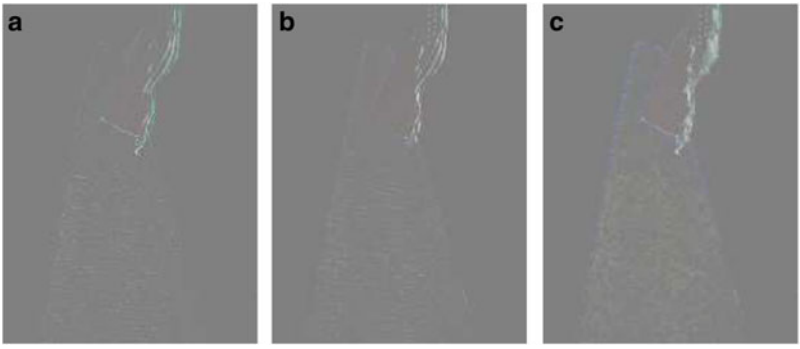


Fig. 20 Matrix-based black, white, and self-dual top-hat applied to an natural image. (**a**) Black top-hat. (**b**) White top-hat. (**c**) Self-dual top-hat

6 Summary and Future Work

The real symmetric 2×2-matrices of a matrix field offer three degrees of freedom, enough to house the three components of many popular color models, such as the RGB- or the HSI-model. The matrix-setting has indeed several advantages over the vector-setting of color images: First the richer algebraic structure, second, the existence of an almost canonical order, the Loewner order. And third, when morphological operations are applied, the interaction of the color channels is ensured, if these are coded in matrix form. This coding was inspired by the close, almost obvious geometric relation between the HC$\tilde{\text{L}}$ bicone, a variant of the HSI-bicone, and an order interval induced by the Loewner order cone. Once

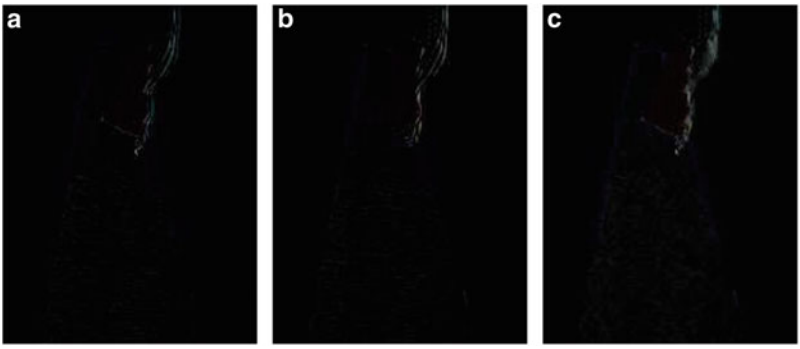


Fig. 21 Component-wise approach: The morphological operations of black, white and self-dual top-hat applied RGB-channel-wise to an natural image. (**a**) Component-wise black top-hat. (**b**) Component-wise white top-hat. (**c**) Component-wise self-dual top-hat

the color image has been rewritten in this manner, the morphological techniques developed for matrix fields in [7] and [8] were applicable and the results of various morphological operations ranging from the elementary dilation and erosion to second order derivatives could be studied.

The experimental results are promising and suffice as a proof-of-concept for this novel approach to color image morphology.

However, in the future we will investigate the usefulness of other color models as well as their embedding into the matrix setting, and we will attempt to develop meaningful notions of top-hats, and other more sophisticated morphological operators for color images.

Acknowledgements The authors would like to thank Sheng Xu, Robert M. Freund, and Jie Sun for the MATLAB source code of the subgradient method.

References

1. Agoston, M.K.: Computer Graphics and Geometric Modeling: Implementation and Algorithms. Springer, London (2005)
2. Aptoula, E., Lefèvre, S.: A comparative study on multivariate mathematical morphology. Pattern Recognit. **40**(11), 2914–2929 (2007)
3. Barnett, V.: The ordering of multivariate data. J. Stat. Soc. **A 139**(3), 318–355 (1976)
4. Borwein, J.M., Lewis, A.S.: Convex Analysis and Nonlinear Optimization. Springer, New York (2000)
5. Burgeth, B., Welk, W.,Feddern, Ch., Weickert, J.: Morphological operations on matrix-valued images. In: Pajdla, T., Matas, J. (eds.) Computer Vision – ECCV 2004, Part IV, Prague. Lecture Notes in Computer Science, vol. 3024, pp. 155–167. Springer, Berlin (2004)
6. Burgeth, B., Bruhn, A., Papenberg, N., Welk, M., Weickert, J.: Mathematical morphology for tensor data induced by the Loewner ordering in higher dimensions. Signal Process. **87**(2), 277–290 (2007)

7. Burgeth, B., Papenberg, N., Bruhn, A., Welk, M., Feddern, C., Weickert, J.: Mathematical morphology based on the loewner ordering for tensor data. In: Ronse, C., Najman, L., Decencière, E. (eds.) Mathematical Morphology: 40 Years On Computational Imaging and Vision, vol. 30, pp. 407–418. Springer, Dordrecht (2005)
8. Burgeth, B., Welk, M., Feddern, C., Weickert, J.: Mathematical morphology on tensor data using the Loewner ordering. In: Weickert, J., Hagen, H. (eds.) Visualization and Processing of Tensor Fields, pp. 357–367. Springer, Berlin (2006)
9. Comer, M.L., Delp, E.J.: Morphological operations for color image processing. J. Electron. Imaging **8**(3), 279–289 (1999)
10. Serra, J.: Anamorphoses and function lattices. In: Dougherty, E.R. (ed.) Mathematical Morphology in Image Processing, pp. 483–523. Marcel Dekker, New York (1993)
11. Gaertner, B.: Smallest enclosing balls of points – fast and robust in **C++**. http://www.inf.ethz.ch/personal/gaertner/miniball.html. Last visited: 03 July 2012
12. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Addison–Wesley, Reading (2002)
13. Goutsias, J., Heijmans, H.J.A.M., Sivakumar, K.: Morphological operators for image sequences. Comput. Vis. Image Underst. **62**, 326–346 (1995)
14. Haralick, R.M.: Digital step edges from zero crossing of second directional derivatives. IEEE Trans. Pattern Anal. Mach. Intell. **6**(1), 58–68 (1984)
15. Heijmans, H.J.A.M.: Morphological Image Operators. Academic, Boston (1994)
16. Huppert, B.: Angewandte Lineare Algebra. de Gruyter, New York (1990)
17. Kimmel, R., Bruckstein, A.M.: Regularized Laplacian zero crossings as optimal edge integrators. Int. J. Comput. Vis. **53**(3), 225–243 (2003)
18. Koenderink, J.J.: Color for the Sciences. MIT, Cambridge (2010)
19. Kramer, H.P., Bruckner, J.B.: Iterations of a non-linear transformation for enhancement of digital images. Pattern Recognit. **7**, 53–58 (1975)
20. Marr, D., Hildreth, E.: Theory of edge detection. Proc. R. Soc. Lond. Ser. B **207**, 187–217 (1980)
21. Matheron, G.: Eléments pour une théorie des milieux poreux. Masson, Paris (1967)
22. Serra, J.: Echantillonnage et estimation des phénomènes de transition minier. Ph.D. thesis, University of Nancy (1967)
23. Serra, J.: Image Analysis and Mathematical Morphology, vol. 1. Academic, London (1982)
24. Serra, J.: Image Analysis and Mathematical Morphology, vol. 2. Academic, London (1988)
25. Soille, P.: Morphological Image Analysis, 2nd edn. Springer, Berlin (2003)
26. van Vliet, L.J., Young, I.T., Beckers, A.L.D.: A nonlinear Laplace operator as edge detector in noisy images. Comput. Vis. Graph. Image Process. **45**(2), 167–195 (1989)
27. Xu, S., Freund, R.M., Sun, J.: Solution methodologies for the smallest enclosing circle problem. High Perform. Comput. Eng. Syst. (HPCES). http://hdl.handle.net/1721.1/4015 (2003)

Sharpening Fibers in Diffusion Weighted MRI via Erosion

Thomas C.J. Dela Haije, Remco Duits, and Chantal M.W. Tax

Abstract In this chapter erosion is generalized to the space of diffusion weighted MRI data. This is done effectively by solving a Hamilton-Jacobi-Bellman (HJB) system (erosion) on the coupled space of three dimensional positions and orientations, embedded as a quotient in the group of three dimensional rigid body motions. The solution to the HJB equations is given by a well-posed morphological convolution. We present two numerical approaches to solve the HJB equations: analytical kernels, and finite differences. Proof of concept is given by showing improved visibility of major fiber bundles in both artificial and human data. Furthermore, the method is shown to significantly improve the output of a probabilistic tractography algorithm used to extract the optic radiation.

Keywords Diffusion weighted MRI • Erosion • Hamilton-Jacobi-Bellman equations • Lie groups • Regularization • Sharpening • Sub-Riemannian • Geometry

1 Introduction

Diffusion Weighted MRI (DW-MRI) is a collection of magnetic resonance imaging techniques used to infer structural information from fibrous tissue such as the brain white matter. DW-MRI locally measures diffusion [46], the random motion of molecules, in one or multiple directions. Because this motion is less constrained along the fiber direction than across, these measurements effectively characterize

T.C.J. Dela Haije (✉) • R. Duits
Imaging Science and Technology Eindhoven (IST/e), Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: T.C.J.Dela.Haije@tue.nl,R.Duits@tue.nl

C.M.W. Tax
Image Sciences Institute (ISI), University Medical Center Utrecht, Utrecht, The Netherlands
e-mail: chantal@isi.uu.nl

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_5,

tissue orientation. A prevalent means to analyze this data is tractography, which tries to join locally registered fiber fragments based on their alignment, producing fiber tracts that potentially represent nerve bundles. A simpler application of DW-MRI is for instance the diagnosis and assessment of stroke, where swelling of the cells appears to hinder diffusion significantly [51].

1.1 Diffusion Weighted MRI

Interpretation of the diffusion measurements is an area of active research. A variety of models have been suggested to extract diffusion features, the most common of which is Diffusion Tensor Imaging (DTI) [2, 4, 34, 36]. In DTI, the measured diffusion in a number of directions $\mathbf{n} \in S^2$ after a time t is assumed to satisfy an anisotropic Gaussian distribution. The space S^2 is the ordinary sphere given by the collection of $\mathbf{x} \in \mathbb{R}^3$ that satisfy $\|\mathbf{x}\| = 1$. A symmetric, second order, three-dimension tensor $D(\mathbf{x})$ is estimated at each spatial point $\mathbf{x} \in \mathbb{R}^3$, such that the local amount of diffusion $p(\mathbf{n}|\mathbf{x})$ in the direction $\mathbf{n}$ is given by [1]

$$p(\mathbf{n}|\mathbf{x}) = \frac{1}{2\pi t^{\frac{3}{2}}\sqrt{|\det D(\mathbf{x})|}} \left(\frac{\mathbf{n}^T (D(\mathbf{x}))^{-1} \mathbf{n}}{t} \right)^{-\frac{3}{2}}. \tag{1}$$

More complicated models such as the diffusion orientation transform [39] use more measurements to reconstruct this $p(\mathbf{n}|\mathbf{x})$, called the Orientation Distribution Function (ODF), with fewer model assumptions [28].

From the measured ODF, which gives the diffusion likelihood in a certain direction *given* a position, we find the likelihood at any combination $(\mathbf{x}, \mathbf{n})$ from

$$p(\mathbf{x}, \mathbf{n}) = p(\mathbf{n}|\mathbf{x})\, p(\mathbf{x}), \tag{2}$$

where $p(\mathbf{x})$ is an a priori distribution on position space only, and $(\mathbf{x}, \mathbf{n}) \mapsto p(\mathbf{x}, \mathbf{n})$ is a spatial distribution of diffusivity profiles. A typical function description of $p(\mathbf{x})$ would be the normalized indicator function $p(\mathbf{x}) = \frac{1_\Omega(\mathbf{x})}{\mu_\Omega}$, where $\Omega \subset \mathbb{R}^3$ is the region of the brain that contains the white matter, and μ is the measure.

Diffusivity profiles are typically visualized by morphed spheres called glyphs. The diffusion likelihood $p(\mathbf{x}, \mathbf{n})$ determines the radius of the sphere at $\mathbf{x}$ in the direction $\mathbf{n}$.

1.2 Enhancement and Tractography

While DW-MRI is unique in its ability to measure the microstructure of the brain in vivo, it has some notable drawbacks. The scanning time, primarily determined by

the number of directions needed to reliably fit a certain model, can be in the order of hours for the more complex scanning protocols [52]. Decreasing the number of scanning directions generally decreases the reliability of a fit. Secondly, the scanning time is usually minimized by utilizing very fast imaging sequences such as the echo-planar imaging sequence. These invariably result in noisy data, resulting in a less reliable signal.

The consequences of this are apparent. While the obtained glyphs will still generally point in the directions of the major fiber bundles at each point, simply looking at the local maximum likelihoods is often not enough to reliably determine these directions. Following the emergence of the tractography methods that are based on these directions, the importance of enhancement techniques in neurological applications pertaining to tractography grew [7–9, 40, 47–49].

1.3 Sharpening

Difficulties still remain though, as enhancement inadvertently lowers the level of detail and contrast. To remedy this, one typically uses sharpening methods that emphasize prominent features by attempting to increase contrast at interesting points in the data, typically edges. Previous attempts to sharpen DW-MRI data sets [16, 25, 26, 29, 38] have primarily focused on angular sharpening, i.e. the sharpening of each glyph individually. Promising results have been obtained, but by only considering angular sharpening one by definition ignores the available spatio-angular information. Secondly, many of the proposed methods are based on deconvolution, a process which is in principle ill-posed.

In this article we outline a well-posed alternative based on grayscale morphology, extending well established works [6, 50, 53] on $\mathbb{R}^3$ to the space of diffusion images. By defining an erosion operator for DW-MRI we can slim fibrous structures in the spatial and the angular domain simultaneously. To improve reliability, we take into account the context: if a glyph and its neighbors are aligned, then the likelihood that the direction of alignment is equal to the direction of the nervous fiber bundle is increased. This is illustrated in Fig. 1. Contextual processes have previously been shown to be useful in enhancements [13, 18, 20, 41–43].

Our framework relates to the general framework of group morphology [44] applied to the group of three dimensional rigid body motions $SE(3)$. In addition to Roerdink [44] we rely on scale-space PDE's whose solutions are given by a morphological convolution with the corresponding Green's function. Furthermore, we extend the morphological group convolutions on $SE(3)$ to morphological convolutions on the quotient $SE(3)/(\{\mathbf{0}\} \times SO(2))$, see Sect. 3.3.1.

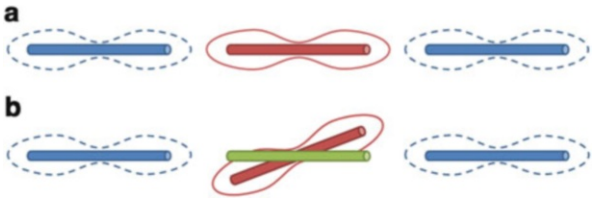


Fig. 1 If we take into account only the local diffusivity profiles, then both of the orientations indicated by the *red bars* would be considered equally correct (**a**). If we take into account the context however, here illustrated by the *blue glyphs*, we will see that in (**b**) the orientation given by the *green bar* seems more likely to be correct than the *red* one. This information can be exploited to improve the results of contextual operators

1.4 Outline of the Article

The article is structured as follows. Section 2 presents a summary of previously published theory, which is used in Sect. 3 to define the erosion operator used to sharpen the DW-MRI data. This section also covers implementation issues. Some practical uses of the erosion operator are lastly given in Sect. 4.

2 Interpreting the Data

We will consider the diffusion images as (twice differentiable) functions on the space of positions and orientations $\mathbb{R}^3 \times S^2$, i.e. $U \in C^2(\mathbb{R}^3 \times S^2, \mathbb{R}^+)$, where $U(\mathbf{x}, \mathbf{n}) = p(\mathbf{x}, \mathbf{n})$ as defined in Eq. (2). To grasp the structure of this space, we must realize that the spaces of positions and orientations are coupled. This can be argued from the concept of alignment, which is impossible to define on the decoupled space, as demonstrated in Fig. 2.

In this section we will show that by embedding the space $\mathbb{R}^3 \times S^2$ as a quotient in the special Euclidean group, we can not only make this idea of a coupled space concrete, but we can also impose in a straightforward way operator invariance under rotations and translations. Once we have clarified the natural group structure embodying $\mathbb{R}^3 \times S^2$, we continue by defining the Euclidean-invariant moving frame of reference and the accompanying legal metrics. These metrics will be used in the next section as the basis for the erosion process. To keep the length of the article at bay, many of the details and proofs are omitted. A more extensive overview of the theory is available [20].

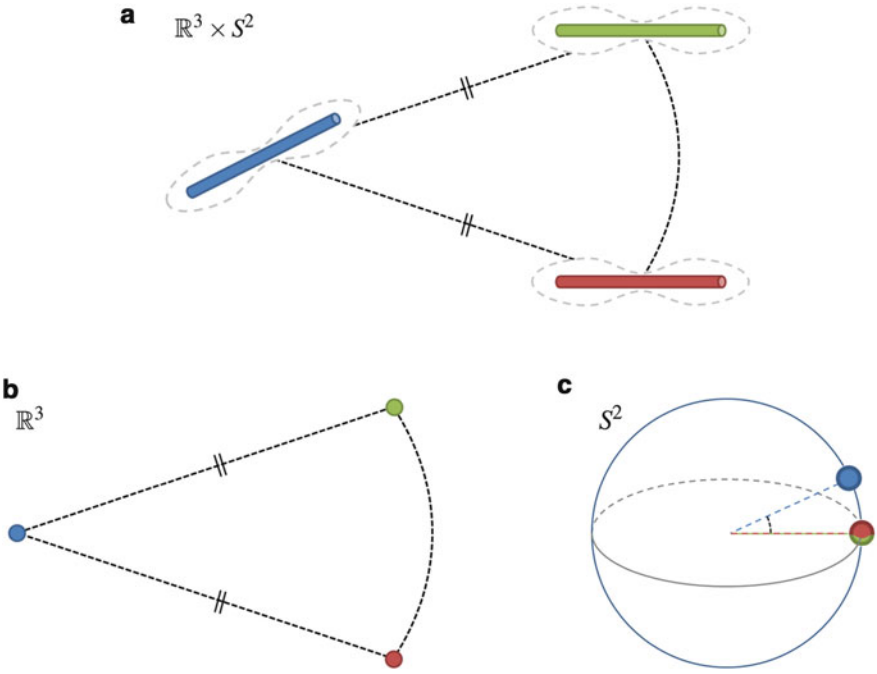


Fig. 2 An illustration in the space $\mathbb{R}^3 \times S^2$ of the coupled nature of alignment. The situation sketched in (**a**) shows three glyphs and their main directions, with the *blue* and the *green glyphs* visually more aligned than the *blue* and the *red glyphs*. The bars that indicate the main directions are elements of the space of positions and orientations $\mathbb{R}^3 \times S^2$. If we now consider projections onto the subspaces $\mathbb{R}^3$ and S^2 *separately* (**b** and **c** respectively) as we do when we consider $\mathbb{R}^3 \times S^2$ as a Cartesian space, we see that we lose the ability to distinguish between the *green* and the *red bars*. This is essentially due to the fact that the spatial and angular distances between the *blue* and the *green bar* are equal to the respective spatial and angular distances between the *blue* and the *red bar*

2.1 The Group Structure and Euclidean-Invariance

As stated above, the manner in which to define this elusive coupling follows naturally from the embedding of $\mathbb{R}^3 \times S^2$ into the special Euclidean group $SE(3)$ of three-dimensional translations and rotations. Elements of $SE(3) = \mathbb{R}^3 \rtimes SO(3)$ are of the form $(\mathbf{x}, Q)$, and the set is endowed with the group operation

$$(\mathbf{x}, Q)(\mathbf{x}', Q') = (Q.\mathbf{x}' + \mathbf{x}, Q.Q').$$

The group of 3D rigid body motion is commonly denoted by $SE(3)$. Intuitively, the group product represents the fact that a concatenation of two rigid body motions ($(\mathbf{x}, Q)$ and $(\mathbf{x}', Q')$) is again a rigid body motion with translation vector $Q.\mathbf{x}' + \mathbf{x}$

and rotation matrix QQ'. The Special Euclidean motion group is the semi-direct product $\mathbb{R}^3 \rtimes SO(3)$ of the translation group $\mathbb{R}^3$ and rotation group $SO(3)$. This is not a direct product as the rotation part affects the product in the position part. Therefore one writes $SE(3) = \mathbb{R}^3 \rtimes SO(3)$ instead of $\mathbb{R}^3 \times SO(3)$. As we will see next, this semi-direct product structure is responsible for a coupling between spatial and angular space, in the space $\mathbb{R}^3 \rtimes S^2$ embedded in $SE(3)$.

The embedding is realized by identifying the elements of $\mathbb{R}^3 \times S^2$ with the elements of $SE(3)$. We do this by defining the intermediary space

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

where we identify $SO(2)$ with rotations around the vector $\mathbf{e}_z$. As a result, elements of $\mathbb{R}^3 \rtimes S^2$ are equivalence classes of $SE(3)$ under the equivalence relation

$$(\mathbf{x}, Q) \sim (\mathbf{x}', Q') \Leftrightarrow \left(\mathbf{x} = \mathbf{x}' \text{ and } \exists_{\alpha \in [0,2\pi)} Q = Q'.Q_{\mathbf{e}_z,\alpha}\right), \tag{3}$$

with $Q_{\mathbf{e}_z,\alpha}$ a rotation by angle α around $\mathbf{e}_z$. Each equivalence class $[(\mathbf{x}, Q)] = \{(\mathbf{x}', Q') \in SE(3) | (\mathbf{x}', Q') \sim (\mathbf{x}, Q)\}$ can be uniquely identified with an element of $\mathbb{R}^3 \times S^2$, using the relation

$$\mathbb{R}^3 \rtimes S^2 \ni [(\mathbf{x}, Q_{\mathbf{n}})] \Longleftrightarrow (\mathbf{x}, \mathbf{n}) \in \mathbb{R}^3 \times S^2, \tag{4}$$

where $Q_{\mathbf{n}}.\mathbf{e}_z \equiv \mathbf{n}$. The explicit relation between elements of $SE(3)$ and $\mathbb{R}^3 \rtimes S^2$ is then given by

$$SE(3) \ni (\mathbf{x}, Q) \longleftrightarrow [(\mathbf{x}, Q)] \in \mathbb{R}^3 \rtimes S^2, \tag{5}$$

where it should be noted that any two distinct elements $(\mathbf{x}, Q_{\mathbf{n}}), (\mathbf{x}', Q'_{\mathbf{n}}) \in [(\mathbf{x}, Q_{\mathbf{n}})]$ are both mapped to the same equivalence class $[(\mathbf{x}, Q_{\mathbf{n}})]$, meaning this relation is not one-to-one. An overview of this embedding is presented in Fig. 3, and details on the parametrization of these spaces can be found in section "Parametrization of the Special Euclidean Group".

2.1.1 Relating Functions on the Group to Functions on the Quotient

To distinguish functions on $SE(3)$ from functions on $\mathbb{R}^3 \rtimes S^2$ and $\mathbb{R}^3 \times S^2$ (which are homeomorphic), the former are denoted $\tilde{U}$ instead of U:

$$\tilde{U} : SE(3) \to \mathbb{R}^+ : (\mathbf{x}, Q) \mapsto \tilde{U}(\mathbf{x}, Q).$$

Following Eqs. (4) and (5), we identify $\tilde{U}$ with a function $U : \mathbb{R}^3 \rtimes S^2 \to \mathbb{R}^+$ according to

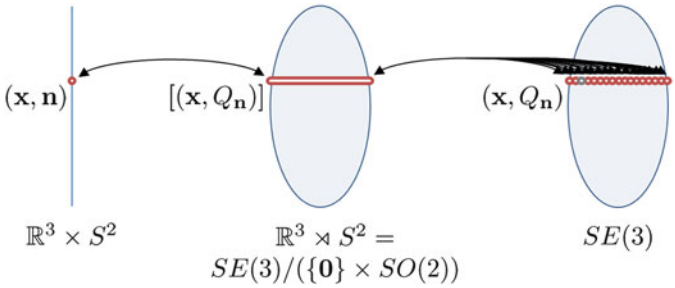


Fig. 3 The quotient $\mathbb{R}^3 \rtimes S^2$ consists of equivalence classes of $SE(3)$ and is homeomorphic to the original space of diffusion images $\mathbb{R}^3 \times S^2$

$$\begin{aligned} U(\mathbf{x}, \mathbf{n}) &= \tilde{U}(\mathbf{x}, Q_{\mathbf{n}}), \\ \tilde{U}(\mathbf{x}, Q) &= U(\mathbf{x}, Q.\mathbf{e}_z), \end{aligned} \tag{6}$$

for all $\mathbf{x} \in \mathbb{R}^3$ and all $Q \in SO(3)$. Recalling Eq. (3), we note that identification is unique if and only if we restrict ourselves to functions on the group with the following invariance property:

$$\tilde{U}(\mathbf{x}, Q) = \tilde{U}(\mathbf{x}, Q.Q_{\mathbf{e}_z,\alpha}), \tag{7}$$

for all $\alpha \in [0, 2\pi)$. If this requirement is not met there is no logical way to choose a function value out of $\{\tilde{U}(g) \mid g \in [(\mathbf{x}, Q_{\mathbf{n}})]\}$ to assign to $U(\mathbf{x}, \mathbf{n})$. The fact that operators on functions on $SE(3)$ have to preserve this property for the resulting functions to be properly defined on the quotient limits the set of possible functions, as will be discussed in the next section.

2.2 *Operator Legality*

If we want to define well-behaving operators on functions on $\mathbb{R}^3 \rtimes S^2$, it is good practice to impose Euclidean-invariance, i.e. invariance of the operator with respect to translations and rotations. Since all operators will first be defined on $SE(3)$, operators are additionally required to be α-invariant, a necessary property that follows from Eq. (3) and that ensures a unique relation between operators on $SE(3)$ and $\mathbb{R}^3 \rtimes S^2$. To define these invariances we define the left- and right group actions

$$\begin{aligned} L_g &: h \mapsto L_g(h) = gh, \\ R_g &: h \mapsto R_g(h) = hg, \end{aligned}$$

and the corresponding operators on a function $\tilde{U}$ by

$$(\mathscr{L}_g \circ \tilde{U})(h) = (\tilde{U} \circ L_g^{-1})(h) = \tilde{U}(g^{-1}h),$$
$$(\mathscr{R}_g \circ \tilde{U})(h) = (\tilde{U} \circ R_g)(h) \;\; = \tilde{U}(hg),$$

for all $g, h \in SE(3)$ and with $\circ$ function composition. $\mathscr{L}$ and $\mathscr{R}$ are respectively called the left- and right-regular representations, as they satisfy

$$\mathscr{L}_g \mathscr{L}_h = \mathscr{L}_{gh} \text{ and } \mathscr{R}_g \mathscr{R}_h = \mathscr{R}_{gh}$$

for all $g, h \in SE(3)$.

2.2.1 α-Invariance

In order to ensure that operators $\tilde{\Phi}$ acting on functions on the group respect the quotient structure depicted in Fig. 3, we must impose invariance under an additional rotation $Q_{\mathbf{e}_z,\alpha}$ from the right. Recall again Eq. (3). This α-invariance is the operator-equivalent of Eq. (7), and requires essentially that any operator $\tilde{\Phi}$ must satisfy

$$\mathscr{R}_g \circ \tilde{\Phi} = \tilde{\Phi} \tag{8}$$

for all $g = (\mathbf{0}, Q_{\mathbf{e}_z,\alpha}) \in (\{\mathbf{0}\} \times SO(2))$.

2.2.2 Left-Invariance

Euclidean-invariance now can be achieved by imposing invariance with respect to either left-regular representations (left-invariance), right-regular representations (right-invariance), or both. It turns out that only $\mathscr{L}_g$ is α-invariant, which along with arguments presented in the PhD thesis of Franken [27, Sect. 7.4.3] shows that for this case only left-invariance should be considered. Hence the invariance-imposed restriction is that every operator $\tilde{\Phi}$ acting on a function $\tilde{U}$ must commute with $\mathscr{L}_g$ for all $g \in SE(3)$:

$$\tilde{\Phi} \circ \mathscr{L}_g = \mathscr{L}_g \circ \tilde{\Phi}. \tag{9}$$

This requirement immediately guarantees Euclidean-invariance of the corresponding operator Φ defined on $\mathbb{R}^3 \rtimes S^2$ by means of Eq. (5).

2.2.3 Legal Operators

Operators on functions defined on $SE(3)$ are called legal if they correspond uniquely to well-defined operators on $\mathbb{R}^3 \rtimes S^2$, meaning they satisfy both Eqs. (8) and (9). Note that any concatenation and linear combination of legal operators is again legal.

2.3 The Moving Frame of Reference

We now define for each pair of a position and an orientation $(\mathbf{x}, \mathbf{n})$, a local frame of vectors aligned both spatially and angularly with $\mathbf{n}$. This enables us to define derivatives in a coordinate frame relative to $(\mathbf{x}, \mathbf{n})$, which ultimately allows us to define erosion at $\mathbf{x}$ towards fibers oriented along $\mathbf{n}$. As we do this for all $(\mathbf{x}, \mathbf{n})$ in the data set $(\mathbf{x}, \mathbf{n}) \mapsto U(\mathbf{x}, \mathbf{n})$, we can erode towards likely fibers present in the data.

As we have identified $\mathbb{R}^3 \times S^2$ with $SE(3)$, this frame can be expressed conveniently on the group $SE(3)$ first. From a geometrical point of view, this boils down to the definition of a moving frame of reference on the manifold $SE(3)$, which is simply a collection of vector fields that span the tangent spaces at each point on the manifold. The erosion operator that we will define in the next section will be expressed in terms of these vector fields.

Remark 1. Vector fields can always be considered as differential operators on locally defined smooth functions [3].

Because we plan to use the moving frame to define an erosion evolution, it is convenient to already incorporate left-invariance. That way any linear combination of these vector fields will be left-invariant as well, meaning we will only have to deal with α-invariance in order to assure legality. The left-invariant vector fields can be generated from any basis of the tangent space at the unity element (the Lie algebra). Here we will adhere to the following basis vectors at the unity element $e = (\mathbf{0}, I)$:

$$A_1 = \partial_x|_e, A_2 = \partial_y|_e, A_3 = \partial_z|_e, A_4 = \partial_{\tilde{\gamma}}|_e, A_5 = \partial_{\tilde{\beta}}|_e, A_6 = \partial_{\tilde{\alpha}}|_e,$$

where $\partial_{x^i}|_g \tilde{U} = \frac{\partial \tilde{U}}{\partial x^i}(g) \equiv \partial_i|_g \tilde{U}$, and where the variables of differentiation come from the second parameterization chart given in section "Parametrization of the Special Euclidean Group". Writing $\mathscr{A}_i|_g$ for the ith basis vector at $g \in SE(3)$, we obtain the left-invariant vector fields through the push-forward of the left group action [10] (Fig. 4),

$$\mathscr{A}_i|_g \tilde{U} = (L_g)_*(A_i)\tilde{U} \equiv A_i(\tilde{U} \circ L_g).$$

The dependence of the tangent vectors (and their duals) on the group element g will from hereon be omitted for readability where possible.

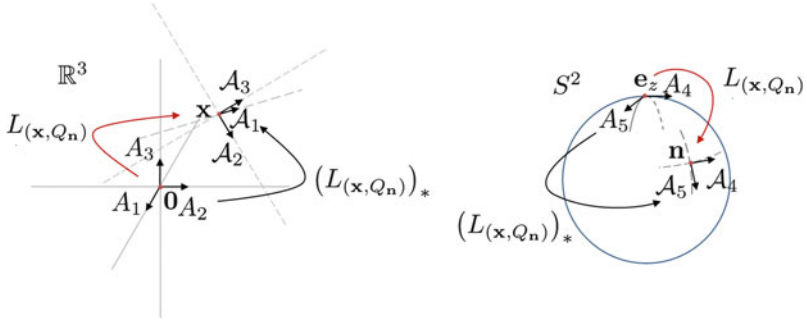


Fig. 4 A schematic representation of the five primary local frame vectors. Movement in the direction of $\mathscr{A}_6$ produces a simultaneous rotation of $\mathscr{A}_1$ and $\mathscr{A}_2$ around $\mathscr{A}_3$, and of $\mathscr{A}_4$ and $\mathscr{A}_5$ around $\mathbf{n} \equiv \mathscr{A}_3$. Since all functions we consider satisfy Eq. (7), $\mathscr{A}_6\tilde{U} = 0$. Because of this $\mathscr{A}_6$ is not illustrated, but it may be envisioned in the figure as the normal to the plane span$\{\mathscr{A}_4, \mathscr{A}_5\}$

Remark 2. Note that since we have defined $Q_\mathbf{n}.\mathbf{e}_z = \mathbf{n}$, the spatial generator $\mathscr{A}_3|_{(\mathbf{x},Q_\mathbf{n})}$ is always aligned with $\mathbf{n}$, i.e. $n^1\partial_x + n^2\partial_y + n^3\partial_z = \mathscr{A}_3|_{(\mathbf{x},\mathbf{n})}$ where $(n^1, n^2, n^3)^T = \mathbf{n}$ and $(x, y, z)^T = \mathbf{x}$.

Remark 3. In both charts, we have that the final angles α and $\tilde{\alpha}$ are in fact redundant, cf. Eq. (7) and the appendix.

The stated definition of the left-invariant vector fields may also be obtained by taking the derivative of the right-regular representation $\mathscr{R}$, cf. [18] which gives the alternative, more manageable expression

$$\mathscr{A}_i|_g\tilde{U} = (\mathrm{d}\mathscr{R}(A_i)\tilde{U})(g) = \lim_{t\downarrow 0}\frac{\tilde{U}(g\mathrm{e}^{tA_i}) - \tilde{U}(g)}{t}. \tag{10}$$

Here the exponential map takes a vector in the tangent space at e to an open subset Ω_e of the manifold around e, i.e. $\exp : T_e(SE(3)) \to \Omega_e$, see Fig. 5. As it can be shown that this maps $T_e(SE(3))$ bijectively onto $SE(3)$, we may take $\Omega_e = SE(3)$. Explicit expressions for the left-invariant vector fields, their duals, the exponential map, and its inverse, can all be found in previous work [21].

The corresponding frames on $\mathbb{R}^3 \times S^2$ are given by Duits et al. [20]

$$\begin{aligned}
\mathscr{A}_1U(\mathbf{x},\mathbf{n}) &= \lim_{h\downarrow 0}\tfrac{U(\mathbf{x}+hQ_\mathbf{n}\mathbf{e}_x,\mathbf{n})-U(\mathbf{x},\mathbf{n})}{h}\\
\mathscr{A}_2U(\mathbf{x},\mathbf{n}) &= \lim_{h\downarrow 0}\tfrac{U(\mathbf{x}+hQ_\mathbf{n}\mathbf{e}_y,\mathbf{n})-U(\mathbf{x},\mathbf{n})}{h}\\
\mathscr{A}_3U(\mathbf{x},\mathbf{n}) &= \lim_{h\downarrow 0}\tfrac{U(\mathbf{x}+hQ_\mathbf{n}\mathbf{e}_z,\mathbf{n})-U(\mathbf{x},\mathbf{n})}{h}\\
\mathscr{A}_4U(\mathbf{x},\mathbf{n}) &= \lim_{h\downarrow 0}\tfrac{U(\mathbf{x},Q_\mathbf{n}Q_{\mathbf{e}_x,h}\mathbf{e}_z)-U(\mathbf{x},\mathbf{n})}{h}\\
\mathscr{A}_5U(\mathbf{x},\mathbf{n}) &= \lim_{h\downarrow 0}\tfrac{U(\mathbf{x},Q_\mathbf{n}Q_{\mathbf{e}_y,h}\mathbf{e}_z)-U(\mathbf{x},\mathbf{n})}{h}
\end{aligned}$$

This definition will be used for the implementation, but we will for now continue with the vector fields $\{\mathscr{A}_i\}_{i=1}^6$ defined on $SE(3)$.

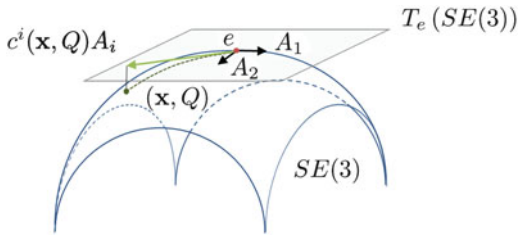


Fig. 5 A simplified diagram of the algebraic structure discussed in this chapter. The bent surface represents $SE(3)$, with the unity element e indicated by the *red dot*. The tangent surface $T_e(SE(3))$ is the set of the Lie algebra, with the basis $\{A_i\}_{i=1}^6$, of which for obvious reasons only the first two are shown. The *green arrow* is a random tangent vector $c^i(\mathbf{x}, Q)A_i$ at the unity element, where $c^i(\mathbf{x}, Q)A_i \xrightarrow{\exp} (\mathbf{x}, Q)$

Remark 4. Note that $\{\mathscr{A}_i\}_{i=1}^5$ are well-defined vector fields on $\mathbb{R}^3 \times S^2$ but *not* on $\mathbb{R}^3 \rtimes S^2$, since $\{\mathscr{A}_i\}_{i\in\{1,2,4,5\}}$ depend on the redundant angle α.

2.4 Legal Metrics

The left-invariant vector fields induce a set of metrics on $SE(3)$. By definition, any metric on $SE(3)$ can be written in the form

$$\mathbf{G}|_{(\mathbf{x},Q)} = \sum_{i,j=1}^{6} g_{ij}(\mathbf{x}, Q)\, \omega^i|_{(\mathbf{x},Q)} \otimes \omega^j|_{(\mathbf{x},Q)}. \tag{11}$$

Here $\omega^i|_{(\mathbf{x},Q)}$ denotes[1] the dual of the ith basis vector of the tangent space at $(\mathbf{x}, Q) \in SE(3)$, with the defining property

$$\omega^i\left(\sum_{j=1}^{6} c^j \mathscr{A}_j\right) = c^i.$$

The coefficients $\{g_{ij}\}_{i,j=1}^6$ represent the relative penalty associated with moving in the corresponding direction $\mathscr{A}_i$, and are subject to the constraint[2]

$$[g_{ij}] = \mathrm{diag}(D^{11}, D^{11}, D^{33}, D^{44}, D^{44}, D^{66})^{-1}, \quad D^{ii} \in \mathbb{R}^+, \tag{12}$$

[1] In previous work we wrote $d\mathscr{A}^i$ for the dual vectors, but to avoid possible confusion with the exterior derivative operator we use ω^i here.

[2] We take here D^{ii} to be constant. This is not strictly necessary, and taking $D^{ii}(\tilde{U})$ can in fact be useful [20].

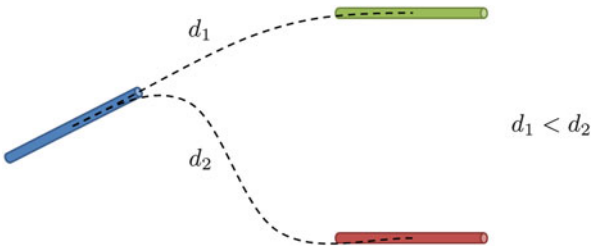


Fig. 6 An example of the use of a sub-Riemannian metric. By prohibiting motion perpendicular to the fiber fragment, we find a distance measure that allows us to distinguish between the aligned and the misaligned glyph of Fig. 2. The *dotted lines* are projections of two (estimated) solutions to Eq. (13) onto $\mathbb{R}^3 \times S^2$

so as to ensure legality [20]. Since any movement along $\mathscr{A}_6$ can be considered null (see Remark 3), we neglect D^{66}.

As an example, assume that $D^{33} = D^{44}$, meaning the distance increases equally with an equal movement in either $\mathscr{A}_3$, $\mathscr{A}_4$, or $\mathscr{A}_5$. Furthermore, we prohibit movement in the directions $\mathscr{A}_1$ and $\mathscr{A}_2$ by taking $D^{11} \downarrow 0$. Note that this implies infinite cost for movement in the plane $\text{span}\{\mathscr{A}_1, \mathscr{A}_2\}$, perpendicular to the spatial propagation direction $\mathscr{A}_3$. This last decision means we are now considering a distance on a sub-Riemannian manifold [5] $B := (SE(3), \ker\{\omega^1\} \cap \ker\{\omega^2\} \cap \ker\{\omega^3\}, \sum_{i=3}^{5} \left(D^{ii}\right)^{-1} \omega^i \otimes \omega^i)$. The distance between two points $g_1, g_2 \in SE(3)$, and the corresponding distance in $\mathbb{R}^3 \times S^2$, is then given by the shortest smooth connecting curve $\bar{\gamma} : [0, L] \to SE(3)$ defined on B (Fig. 6):

$$\mathrm{d}(g_1, g_2) = \inf_{\substack{\bar{\gamma} \in C^\infty([0,L],B),\\ \bar{\gamma}(0) = g_1,\ \bar{\gamma}(L) = g_2,\\ \dot{\bar{\gamma}}^1(s) = \dot{\bar{\gamma}}^2(s) = \dot{\bar{\gamma}}^6(s) = 0}} \int_0^L \sqrt{\frac{1}{D^{33}} \dot{\bar{\gamma}}^3(s)^2 + \frac{1}{D^{44}} (\dot{\bar{\gamma}}^4(s)^2 + \dot{\bar{\gamma}}^5(s)^2)}\, ds. \tag{13}$$

While this illustrates the possibilities of the presented theory, it is an entirely different optimal control problem [15] than the one needed for sharpening. Below we will use another configuration of parameters, giving rise to a different sub-Riemannian manifold and metric tensor that when used to generate an evolution equation, results in the $\mathbb{R}^3 \times S^2$-analogue of typical spatial erosion evolutions.

2.5 Overview

In summary, we have argued that DW-MRI data sets should be considered functions on the coupled space $\mathbb{R}^3 \rtimes S^2$, whose natural group structure follows from embedding the space in the group of three dimensional rigid motions $SE(3)$. By

posing that operators on DW-MRI data should be invariant under rotations and translation, and need to satisfy α-invariance, Eq. (8), we defined a requirement for operators to be 'legal'. We then proceeded to describe a left-invariant moving frame of reference on $SE(3)$, which allows us to look at the data as if attached to a local fiber fragment. This moving frame is then used to describe the range of possible legal metrics, which we can use in the next section to define the erosion operator.

3 Erosion

At an elementary level, morphological operations on $\mathbb{R}^n$ can be considered solutions to a specific class of evolution equations [50]. Writing $f : \mathbb{R}^n \to \mathbb{R}^+$ for the gray value image and $m : \mathbb{R}^n \to \mathbb{R}^+$ for the structuring element, the morphological convolution $h : \mathbb{R}^n \to \mathbb{R}^+$ that solves the erosion equation (which depends on the structuring element m), is given by

$$h(\mathbf{x}) = (f \ominus m)(\mathbf{x}) := \inf_{\mathbf{y}\in\mathbb{R}} [f(\mathbf{y}) + m(-\mathbf{y} + \mathbf{x})]. \tag{14}$$

In case the structuring element satisfies the semi-group property [6, 53], the PDE satisfied by h dictates a morphological scale-space. If for example m is a quadratic structuring element, i.e. of the form $m_t(\mathbf{x}) = \frac{2\eta-1}{2\eta}\left(\frac{\|\mathbf{x}\|^{2\eta}}{t}\right)^{\frac{1}{2\eta-1}}$, where $\eta \in \left(\frac{1}{2}, 1\right]$ and $t \in \mathbb{R}^+$, then the size t of the structuring element parameterizes a morphological scale space [6, 31, 53] dictated by the evolution equation

$$\begin{cases} \partial_t h(\mathbf{x}; t) = -\frac{1}{2\eta}\|\nabla h(\mathbf{x}; t)\|^{2\eta}, \\ h(\mathbf{x}; 0) = f(\mathbf{x}). \end{cases}$$

In generalizing these results to $SE(3)$ we start from the premise that erosion is described by one of the legal scale spaces that can be defined on that group. Similar scale-spaces can be constructed [17] by employing the previously defined general metric. This yields for $g \in SE(3)$ and $t \geq 0$

$$\begin{cases} \partial_t \tilde{W}(g; t) = \pm\frac{1}{2\eta}\left[\mathbf{G}|_g^{-1}\left(d\tilde{W}(g; t), d\tilde{W}(g; t)\right)\right]^{\eta}, \\ \tilde{W}(g; 0) = \tilde{U}(g), \end{cases} \tag{15}$$

where $\tilde{W}(.; t)$ is the function on the Lie group $SE(3)$, initially given by the original data $\tilde{U} : SE(3) \to \mathbb{R}^+$, $d\tilde{W}(g; t) = \sum_{i=1}^{6} \mathscr{A}_i \tilde{W}(g; t)\omega^i$ is the gradient of $\tilde{W}$, and

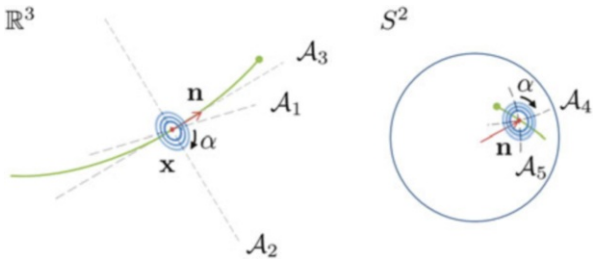


Fig. 7 A schematic showing the rationale behind the choice $D^{33} = 0$ when considering erosion. Since $\mathscr{A}_3$ is oriented along the fiber fragment, one prefers to erode *only* perpendicular to this direction, as indicated by the *blue circles*. Note that the depicted fiber represents *any* potential fiber passing through **x** with orientation **n**

G is the left-invariant metric tensor given in Eq. (11) with the coefficients $D^{ii} \in \mathbb{R}^+$ as in Eq. (12). Different choices[3] for the coefficients result in different scale spaces.

3.1 Erosion Towards Fibers

Proper choices for D^{11}, D^{33} and D^{44} are easily 'guessed' from their relation to the moving frame of reference. As illustrated in Fig. 7, erosion should transport data surrounding a fiber (i.e. perpendicular to $\mathscr{A}_3$) towards it, both spatially and angularly. This means that we need to take $D^{33} = 0$ while D^{11} and D^{44} are still free, and that we should look at the minus case of Eq. (15). The resulting differential equation is the Hamilton-Jacobi-Bellman equation

$$\begin{cases} \partial_t \tilde{W}(g;t) = -\frac{1}{2\eta}\left[D^{11}\left((\mathscr{A}_1\tilde{W}(g;t))^2 + (\mathscr{A}_2\tilde{W}(g;t))^2\right) + \right. \\ \qquad\qquad \left. D^{44}\left((\mathscr{A}_4\tilde{W}(g;t))^2 + (\mathscr{A}_5\tilde{W}(g;t))^2\right)\right]^{\eta}, \\ \tilde{W}(g;0) = \tilde{U}(g), \end{cases} \tag{16}$$

with again $\eta \in \left(\frac{1}{2}, 1\right]$. The proof showing that this evolution equation can be solved by a morphological convolution,[4] and can thus be qualified as an erosion, is fairly technical and will not be given here, but can be found online [20].

[3]The values allowed for the coefficients are subject to the Hörmander requirement [30] which guarantees smooth non-singular scale spaces. Proofs that the evolutions presented here satisfy this condition are available [20].

[4]The morphological convolution is in fact the viscosity solution to the morphological scale space, similar to the same problem on $\mathbb{R}^n$ [11,24] and on the Heisenberg group $H(n)$ [35].

In complete analogy to the problem in $\mathbb{R}^n$, there exists a structuring element, or kernel, $\tilde{k}_t^{D^{11},D^{44}} : SE(3) \to \mathbb{R}^+$, such that

$$\tilde{W}(g;t) = (\tilde{U} \ominus_{SE(3)} \tilde{k}_t^{D^{11},D^{44}})(g),$$

with $\ominus_{SE(3)}$ denoting the $SE(3)$-counterpart to the erosion operator defined in Eq. (14), given by

$$(\tilde{U} \ominus_{SE(3)} \tilde{k}_t^{D^{11},D^{44}})(g) := \inf_{h \in SE(3)} \left[\tilde{U}(h) + \tilde{k}_t^{D^{11},D^{44}}(h^{-1}g) \right], \tag{17}$$

which fits in the framework of group morphology [44]. As shown in other work [6, 19], $\tilde{k}_t^{D^{11},D^{44}}$ is the morphological Green's function, which may be approximated by

$$\tilde{k}_t^{D^{11},D^{44}}(g) \approx \frac{2\eta - 1}{2\eta} \frac{\left(C^2 \sqrt{\frac{(c^3)^2}{D^{11}D^{44}} + \frac{(c^6)^2}{D^{44}D^{44}} + \left(\frac{(c^1)^2+(c^2)^2}{D^{11}} + \frac{(c^4)^2+(c^5)^2}{D^{44}} \right)^2} \right)^{\frac{\eta}{2\eta-1}}}{t^{\frac{1}{2\eta-1}}}, \tag{18}$$

where we use short-hand notation $\{c^i\}_{i=1}^6 = \{c^i(g)\}_{i=1}^6$ coming from the logarithm on $SE(3)$, cf. Fig. 5, of which explicit expression can be found in section "The Logarithmic Map". The constant $C \in (0, 2]$ comes along with the Heisenberg approximation technique [22, 37] used in deriving the expression, and can be considered a simple reparameterization of t. A direct consequence of the fact that the erosion operator on $SE(3)$ is a solution to Eq. (16), is that the morphological convolution given satisfies the semi-group property, i.e.

$$\left((\tilde{U} \ominus_{SE(3)} \tilde{k}_s^{D^{11},D^{44}}) \ominus_{SE(3)} \tilde{k}_t^{D^{11},D^{44}} \right)(g) = (\tilde{U} \ominus_{SE(3)} \tilde{k}_{s+t}^{D^{11},D^{44}})(g)$$

A fairly basic example of erosion is shown in Fig. 8b, where a column of aligned glyphs is surrounded by glyphs with random orientations. The operation significantly decreases the function value for misaligned fiber fragments.

3.2 Minimum Reduction

One property the DW-MRI erosions inherited from the regular $\mathbb{R}^n$ morphological operations, is that bounds on the function given by the extrema are retained. Erosion decreases values that lie near a minimum, but as follows from the fact that the solutions are given by a morphological convolution, they will always remain between the global minimum and maximum. While this guarantees well-posedness and stability with respect to the $\mathbb{L}^\infty$-norm (in contrast to for instance deconvolution

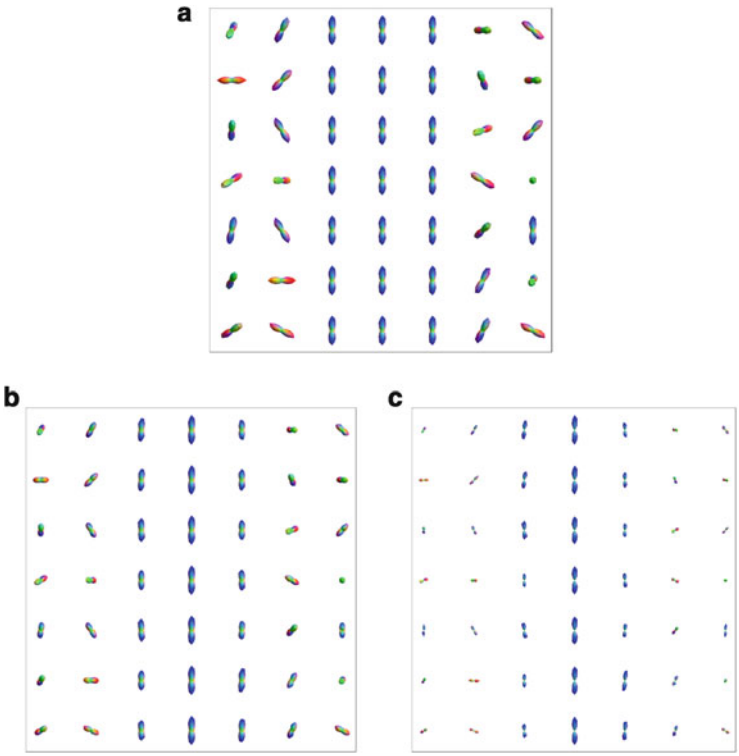


Fig. 8 A basic example of erosion with manually selected parameters ($\Delta t = 0.02$, $D^{11} = 0.5$ and $D^{44} = 0.02$). (**a**) A slice of an artificial data set containing a number of aligned glyphs, surrounded by randomly oriented glyphs. (**b**) The data from (**a**) after erosion, with $t = 3$ and $\eta = 0.75$. (**c**) The data from (**a**) after min-normalization and erosion, with $t = 4$ and $\eta = 0.65$

operators), the drawback of this is that we can not decrease minima even if they seem to be misrepresentative of the local structure. We have previously attempted to resolve this by min-normalizing the data before eroding [14,20]. min-normalization $\nu_{\min}$ of a data set $\tilde{U} : SE(3) \to \mathbb{R}^+$ is defined as

$$\nu_{\min}(\tilde{U})(\mathbf{x}, \mathbf{n}) = \tilde{U}(\mathbf{x}, \mathbf{n}) - \min\{\tilde{U}(\mathbf{x}, \mathbf{n}') | \mathbf{n}' \in S^2\}. \tag{19}$$

The effect of this operation on subsequent erosions is depicted in Fig. 8c. The normalization essentially increases the contrast per glyph, effectively sharpening the data angularly.

A more prevalent approach to angular sharpening in the literature is the $(I - a\Delta_{LB})$ operator, which subtracts the result of the Laplace-Beltrami operator Δ_{LB} scaled by $a \in \mathbb{R}^+$ from the original data:

$$v_{LB}(\tilde{U})(\mathbf{x},\mathbf{n}) = \tilde{U}(\mathbf{x},\mathbf{n}) - a(\Delta_{LB}\tilde{U})(\mathbf{x},\mathbf{n}). \tag{20}$$

In our framework, the Laplace-Beltrami operator is given by

$$\Delta_{LB} = (\mathscr{A}_4)^2 + (\mathscr{A}_5)^2 + (\mathscr{A}_6)^2,$$

where the last term can of course be neglected as we have $(\mathscr{A}_6)^2\tilde{U} = 0$, recall Remark 3.

Remark 5. The Laplace-Beltrami operator is a legal operator, as the α dependence of $\mathscr{A}_4$ and $\mathscr{A}_5$ cancels out in Eq. (3.2).

3.3 Approximating Solutions

We have considered two different approaches to implement erosions for DW-MRI data. The first one is based on the group-convolution with the approximation of the Green's function, Eq. (18), while the second one directly discretizes the Hamilton-Jacobi-Bellman equation that governs erosion, Eq. (16). Both the implementations are included in a Mathematica package available for academic purposes at www.bmia.bmt.tue.nl/people/RDuits/DWIpackage/HARDIAlgorithms.zip. Details of the implementation are briefly outlined below.

3.3.1 Convolutions

In order to approximate the solutions to the erosion equation, we can choose to compute the morphological convolution of Eq. (17) using a discretized version of the erosion kernel. Since we know that due to α-invariance the kernel will be independent of any variation in the redundant angle $\tilde{\alpha}$ (or α in the first chart, section "Parametrization of the Special Euclidean Group"), we can reduce the convolution on $SE(3)$ to

$$(k_t^{D^{11},D^{44}} \ominus_{\mathbb{R}^3\times S^2} U)(g) := \inf_{h\in\mathbb{R}^3\times S^2}\left[U(h) + k_t^{D^{11},D^{44}}(h^{-1}g)\right].$$

where we define the kernel $k_t^{D^{11},D^{44}} : \mathbb{R}^3 \times S^2 \to \mathbb{R}^+$ in terms of $\tilde{k}_t^{D^{11},D^{44}}$ by

$$k_t^{D^{11},D^{44}}(\mathbf{x},\mathbf{n}) := \tilde{k}_t^{D^{11},D^{44}}(\mathbf{x}, Q_{\mathbf{n}}),$$

for any $Q_{\mathbf{n}}$ such that $Q_{\mathbf{n}}.\mathbf{e}_z = \mathbf{n}$.

3.3.2 Discretizing the Problem

Alternatively we can discretize the scale-space equation directly. The initial condition $\tilde{W}(g;0) = \tilde{U}(g)$ of course remains, and the left-hand side of the differential equation becomes

$$\partial_t \tilde{W}(g;t) = \lim_{dt\to 0} \frac{\tilde{W}(g;t+dt) - \tilde{W}(g;t)}{dt} \approx \frac{\tilde{W}(g;t+\Delta t) - \tilde{W}(g;t)}{\Delta t}$$

with Δt small, resulting in the incremental algorithm

$$\tilde{W}(g;t+\Delta t) = \tilde{W}(g;t) \pm \Delta t \frac{1}{2\eta} \left[\mathbf{G}|_g^{-1} \left(d\tilde{W}(g;t), d\tilde{W}(g;t) \right) \right]^{\eta}.$$

In the specific case we are considering here (see Sect. 3.1), we thus get

$$\begin{cases} \tilde{W}(g;t+\Delta t) = \tilde{W}(g;t) - \Delta t \frac{1}{2\eta} \left[D^{11} \left((\mathscr{A}_1 \tilde{W}(g;t))^2 + (\mathscr{A}_2 \tilde{W}(g;t))^2 \right) + \right. \\ \qquad\qquad \left. D^{44} \left((\mathscr{A}_4 \tilde{W}(g;t))^2 + (\mathscr{A}_5 \tilde{W}(g;t))^2 \right) \right]^{\eta}, \\ \tilde{W}(g;0) = \tilde{U}(g), \end{cases}$$

for $g \in SE(3)$, and where we still need to discretize the derivative $\mathscr{A}_i$. At this point we return once more to the space $\mathbb{R}^3 \times S^2$ by means of the identification in Eq. (6):

$$\begin{cases} W(\mathbf{x},\mathbf{n};t+\Delta t) = W(\mathbf{x},\mathbf{n};t) - \Delta t \frac{1}{2\eta} \left[D^{11} \left((\mathscr{A}_1 W(\mathbf{x},\mathbf{n};t))^2 + (\mathscr{A}_2 W(\mathbf{x},\mathbf{n};t))^2 \right) + \right. \\ \qquad\qquad \left. D^{44} \left((\mathscr{A}_4 W(\mathbf{x},\mathbf{n};t))^2 + (\mathscr{A}_5 W(\mathbf{x},\mathbf{n};t))^2 \right) \right]^{\eta}, \\ W(\mathbf{x},\mathbf{n};0) = U(\mathbf{x},\mathbf{n}), \end{cases}$$

We can follow the exact same procedure as before to get the stencils given in section "Finite Difference Schemes".

Remark 6. The spatial derivatives are in fact calculated using an upwind-biased finite differences scheme, see the appendix.

3.3.3 Two Implementations

We distinguished between two implementations for morphological scale spaces:

1. Erosion via convolutions with the approximations of the morphological Green's functions given by Eq. (18).
2. Erosion via left-invariant finite-difference (upwind) schemes, of which the details are provided in section "Finite Difference Schemes".

As explained in detail in Sect. 2.1, the non-commutative nature of $SE(3)$ leads to a natural coupling in the space $\mathbb{R}^3 \rtimes S^2$ of positions and orientations. As a result,

neither implementation is separable in a spatial and angular part, excluding the trivially separable cases where $D^{44}D^{11} = 0$.

Generally speaking, the advantages of the kernel implementations are:

- They allow fast parallel algorithms via lookup tables and precomputed Green's functions, similar to implementations [43] of linear $\mathbb{R}^3 \rtimes S^2$-convolutions [18].
- They are unconditionally stable and directly related to viscosity solutions, cf. [20].
- They involve less interpolation.

The advantages of the finite difference schemes are:

- They are much more flexible towards data adaptive extensions, cf. [14].
- They use efficient (short) stencils of interpolated finite differences, cf. [14].
- Do not involve analytic asymptotical approximations.

The implementations are complementary; the finite differences can be used for accurate precomputation of the Green's functions used in the convolution schemes. Both implementations are included in the *Mathematica* package available (for academic purposes only) at www.bmia.bmt.tue.nl/people/RDuits/DWIpackage/HARDIAlgorithms.zip.

Remark 7. All of the erosions shown in this article have been obtained using the finite differences approach. Note that apart from the scaling factor C of the time, the two methods produce approximately the same results.

4 Preprocessing with Erosion

Qualitative improvement by erosion has been shown before [20], and is reiterated in Fig. 9c. More interesting is the potential benefit of erosion for contextual operators. Figure 9 also shows the use of erosion as a processing step preceding hypo-elliptic diffusion [20].

As stated in the introduction, the most important contextual operators are tractography operators. Probabilistic tractography algorithms cope with data uncertainty by defining a distribution of fiber directions at each position [32, 33], and generate a large amount of pathways by sampling the directions around peaks of this distribution. Sharp and accurate glyphs that are aligned with their context will result in more reliable tractography. Alignment of glyphs is improved here by contour enhancement processes [18]. Since these processes generally propagate oriented particles in too many directions, an extra sharpening step is desired [20].

A challenging fiber bundle to extract is the optic radiation. This structure is part of the visual system, and connects the Lateral Geniculate Nucleus (LGN) to the primary visual cortex $V1$. The optic radiation has a curved anterior extent called Meyer's loop, which makes the fiber bundle difficult to reconstruct reliably. Other fiber bundles in its vicinity further complicate delineation. Direct application of

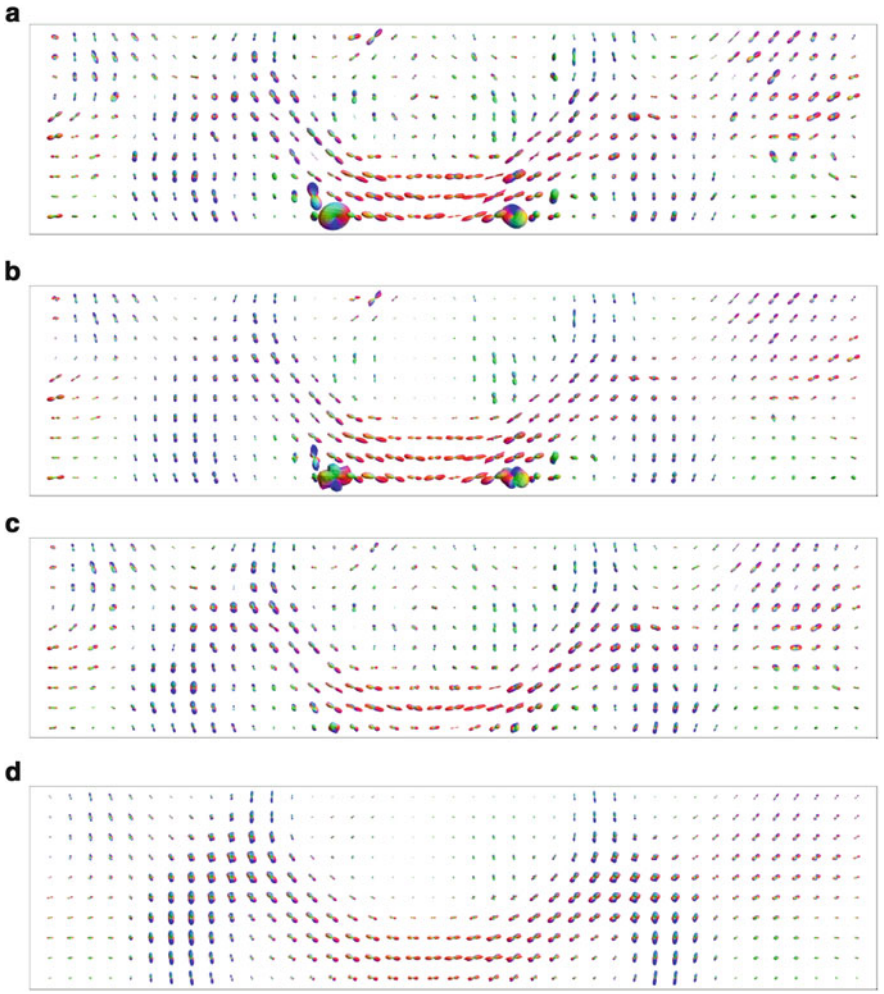


Fig. 9 A DW-MRI data set showing the corpus callosum and the corona radiata, before and after enhancement. Note especially the significant improvement of the enhancement when preceded by erosion. The enhancement was done using hypo-elliptic diffusion [20] with $t = 3$, $\Delta t = 0.01$, $D^{33} = 1$, $D^{44} = 0.002$ and $K = 0.05$. Erosion used $t = 3$, $\Delta t = 0.1$, $\eta = 0.75$, $D^{11} = 1$ and $D^{44} = 0.002$. All visualized data sets were min-normalized using Eq. (19). (**a**) A slice of the min-normalized original data. (**b**) The data from (**a**) after diffusion. (**c**) The data from (**a**) after erosion. (**d**) The data from (**a**) after erosion and diffusion

the reconstruction of this bundle lies in the context of neurosurgical planning for temporal lobe epilepsy. Meyer's loop is often located close to the area that causes epilepsy in these patients, and is disrupted during surgery. This can lead to visual loss of up to a quarter of the visual field.

To show the value of erosion as a preprocessing step for contextual enhancement and tractography, we evaluated the reliability of pathways resulting from a probabilistic tractography algorithm [45] both before and after erosion. The tractography algorithm generates 10^4 tracts based on DTI data[5] using a bootstrapping procedure [23]. We score each pathway $\gamma : [0, L_\gamma] \to \mathbb{R}^3$ according to how well they fit the underlying data, obtained by evaluating U along the tangent vectors $\dot{\gamma} = \frac{d}{ds}\gamma$ of γ [49]:

$$\mathscr{E}_U(\gamma) = \frac{1}{L_\gamma} \int_0^{L_\gamma} \log \left[\frac{U\left(\gamma(s), \dot{\gamma}(s)\right)}{\max U} \right] ds, \tag{21}$$

where L_γ is the length of the pathway γ, and s denotes arc length such that $\dot{\gamma}(s) \in S^2$, i.e. $\|\dot{\gamma}(s)\| = 1$. The data sets are divided by their global maximum for the sake of comparison. The initial cost function U was obtained directly from the diffusion tensors according to Eqs. (1) and (2), where $p(\mathbf{x}) = 1_\Omega(\mathbf{x}) \frac{\sqrt{\det(D(\mathbf{x}))}}{\int_\Omega \sqrt{\det(D(\mathbf{x}))}\, d\mathbf{x}}$ is proportional to the volume of the glyphs. The indicator function 1_Ω is a white matter mask obtained by a fractional anisotropy threshold. The entire scoring pipeline is depicted in Fig. 10.

Figure 11 visualizes the 30, 3 and 0.3 % highest scoring fibers according to Eq. (21) based on the unprocessed data U, the data U pre-processed using only linear hypo-elliptic diffusion, and the data U after both erosion and diffusion. Before any processing, the data is min-normalized.

The method is considered adequate when two conditions are satisfied: on the one hand, the whole optic radiation should be visible (few false negatives), and on the other hand all auxiliary fibers that are not clearly part of the optic radiation should be removed (few false positives). Only when a critical percentage of highest scoring fibers exists that satisfies both conditions, the method is able to give a reliable reconstruction of the optic radiation. We see that only in the case that erosions are included as a preprocessing step, such a critical percentage can be found. In Fig. 11c, we satisfy the two conditions at 28.3 %. There a large portion of the additional fibers falsely identified as part of the optic radiation (false positives) in Fig. 11a, b are removed by the erosion step, which decreases the likelihood of the tracking algorithm finding fibers that deviate strongly from the main fiber orientation.

If we replace the min-normalization with the more common $(I - a\Delta_{LB})$ operator, we can again find a critical percentage at which the method can be considered adequate, though this percentage lies a lot lower than before, at 0.4 %. See also Fig. 12.

[5] Kindly provided by the Kempenhaeghe Epilepsy Center in Heeze, the Netherlands.

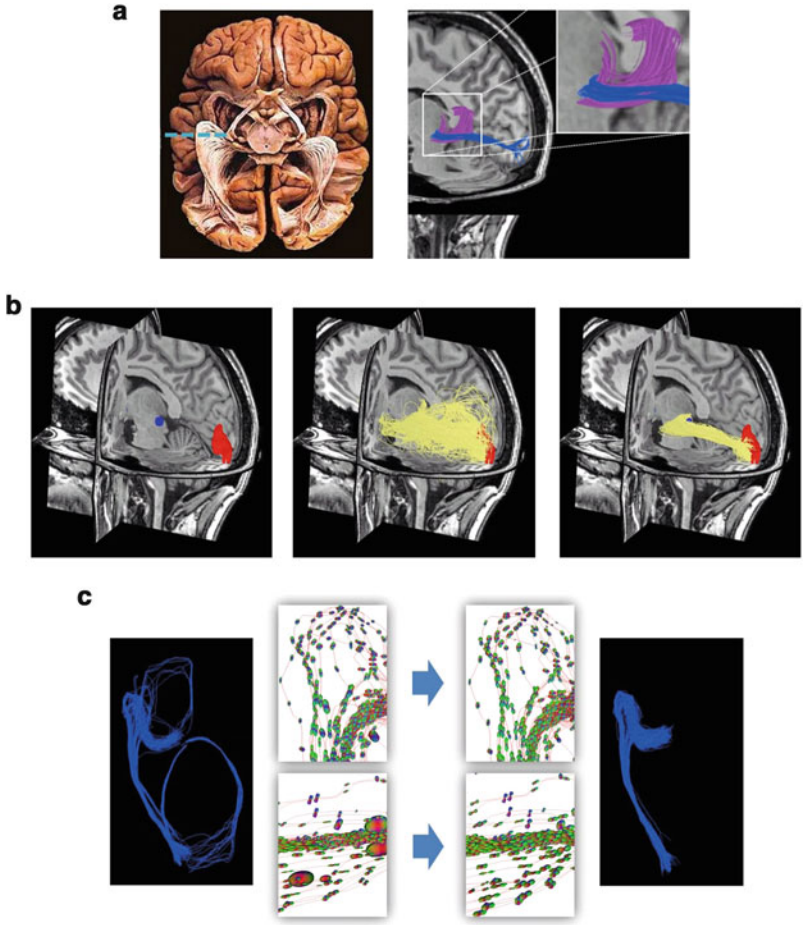


Fig. 10 The scoring pipeline. (**a**) In treating temporal lobe epilepsy, surgeons need to avoid the optic radiation to preserve the patient's vision. Tracking the optic radiation is challenging due to nearby and crossing fibers. (**b**) The optic radiation connects the LGN (blue) and V1 (red). The location of V1 is derived from functional MRI. Probabilistic tractography algorithms generate a large number of tracts between the visual cortex $V1$ and the LGN. These tracts are scored to filter out the optic radiation. (**c**) Enhancement: even with state of the art scoring, the tractography produces a lot of anatomically implausible tracts. Preprocessing the DW-MRI data with erosion and contextual enhancement greatly improves the segmentation of the optic radiation

5 Conclusion

In this paper we have demonstrated the benefits of erosion (well-posed sharpening) as a pre-processing step of contextual processing of Diffusion Weighted MRI. To this end, a Euclidean-invariant erosion evolution was defined on the space $\mathbb{R}^3 \rtimes S^2$

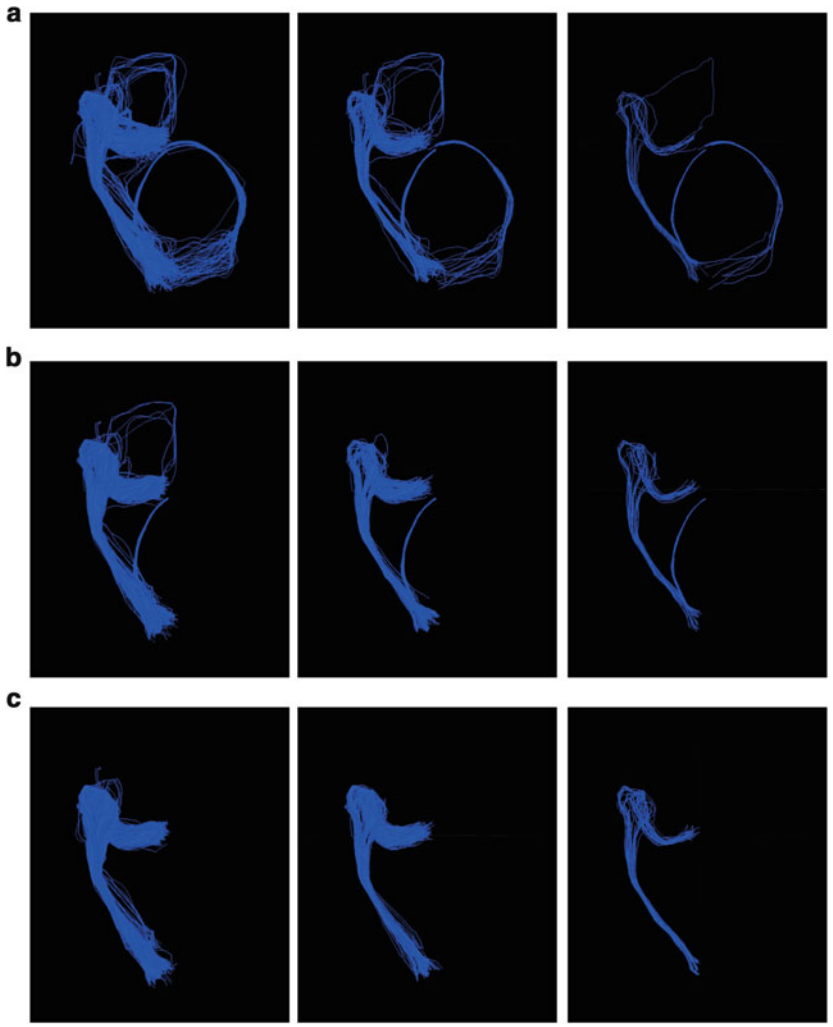


Fig. 11 Tracts obtained from a DTI data set, scored according to differently processed data sets. The hypo-elliptic diffusion has the following parameter settings: $t = 3$, $\Delta t = 0.01$, $D^{33} = 1$ and $D^{44} = 0.01$. Erosion has: $t = 3$, $\Delta t = 0.1$, $\eta = 0.75$, $D^{11} = 1$ and $D^{44} = 0.02$. (**a**) Tracts based on the original data. (**b**) Tracts based on the data from (**a**), following min-normalization and enhancement (hypo-elliptic diffusion). (**c**) Tracts based on the data from (**a**), after min-normalization, erosion, and enhancement

embedded as a quotient in $SE(3)$. The final erosion operator is the mapping that takes the initial condition of this evolution, generally the diffusion weighted image, to the (viscosity) solution with fixed time $t > 0$ of the evolution equation. These erosions satisfy the semi-group property.

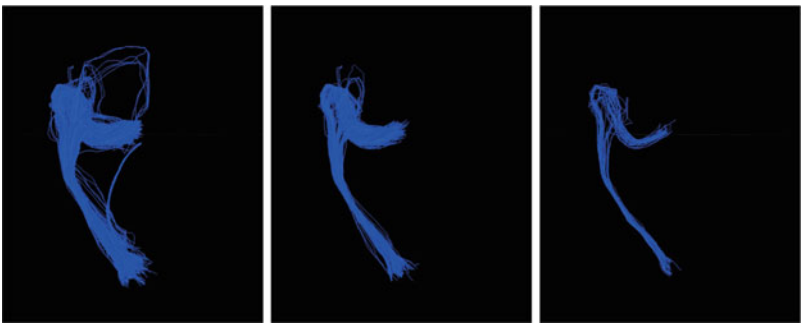


Fig. 12 Tracts obtained in the same way as in Fig. 11c, though the min-normalization is replaced by the minimum reduction technique of Eq. (20) with $a = 0.3$

We have presented two complementary numerical algorithms to compute the erosion operator, an analytical kernel implementation and a finite differences scheme, each with its own advantages.

We have shown the benefits of including erosions in contextual enhancement (via hypo-elliptic diffusion [20]) of DW-MRI. The erosion operator has been shown to visually sharpen the corpus callosum (the major fiber bundle connecting the two hemispheres) and corona radiata which crosses this major fiber bundle radially. See Fig. 9. Finally, we have shown that inclusion of erosions in pre-processing improves subsequent tracking of the optic radiation fibers in the brain, which is relevant for the planning of neurosurgery for epilepsy treatment [49], as shown in Fig. 11.

Parametrization of the Special Euclidean Group

Parametrization of the spatial part of $SE(3)$ is usually done by Cartesian coordinates and for the rotation part $SO(3)$ of $SE(3) = \mathbb{R}^3 \rtimes SO(3)$ we use two charts. Firstly, we use the standard Euler angle parametrization of the sphere given by

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

with $\alpha \in [0, 2\pi)$, $\beta \in [0, \pi]$, $\gamma \in [0, 2\pi)$ and with $R_{\mathbf{v},\phi}$ a counter-clockwise rotation ϕ around the vector $\mathbf{v}$. The rotation axes are depicted in Fig. 13 obtained by applying the rotation to the unit vector oriented along the z-axis. A point $\mathbf{p} \in S^2$ can be identified with all rotations of $\mathbf{e}_z$ such that

$$\mathbf{p} = R_{\mathbf{p}} \mathbf{e}_z \equiv \mathbf{n}(\beta, \gamma), \tag{22}$$

with $R_{\mathbf{p}}$ any R that rotates $\mathbf{e}_z$ to $\mathbf{p}$ and where $\mathbf{n}$ is the parametrization. The parametrization is however ambiguous since

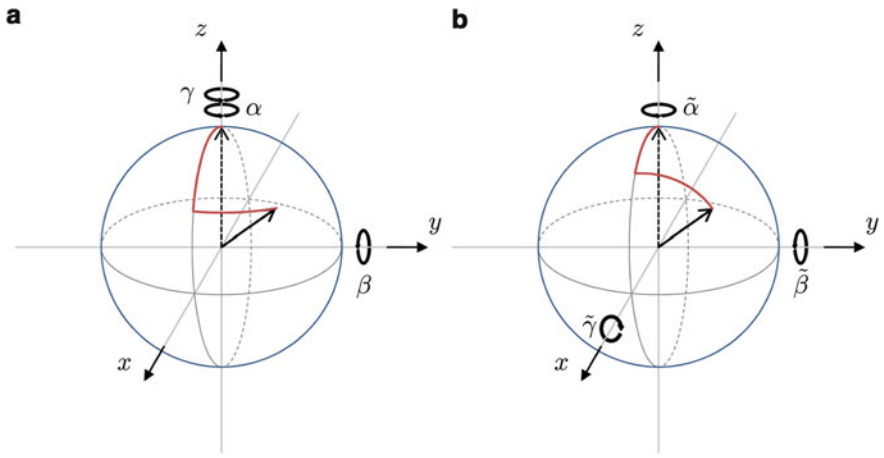


Fig. 13 The two different Euler parameterizations of $SO(3)$ and $S^2 \equiv SO(3)/SO(2)$. (**a**) *ZYZ* parametrization. (**b**) *ZYX* parametrization

$$R_{\mathbf{e}_z,\gamma} R_{\mathbf{e}_y,0} R_{\mathbf{e}_z,\alpha} = R_{\mathbf{e}_z,\gamma+\delta} R_{\mathbf{e}_y,0} R_{\mathbf{e}_z,\alpha-\delta},$$
$$R_{\mathbf{e}_z,\gamma} R_{\mathbf{e}_y,\pi} R_{\mathbf{e}_z,\alpha} = R_{\mathbf{e}_z,\gamma+\delta} R_{\mathbf{e}_y,\pi} R_{\mathbf{e}_z,\alpha-\delta},$$

for all $\delta \in [0, 2\pi)$, and where $R_{\mathbf{e}_z,\gamma} R_{\mathbf{e}_y,0} R_{\mathbf{e}_z,\alpha}$ and $R_{\mathbf{e}_z,\gamma} R_{\mathbf{e}_y,\pi} R_{\mathbf{e}_z,\alpha}$. We consider a second chart

$$\tilde{R} = R_{\mathbf{e}_x,\tilde{\gamma}} R_{\mathbf{e}_y,\tilde{\beta}} R_{\mathbf{e}_z,\tilde{\alpha}},$$

with $\tilde{\beta} \in (-\pi, \pi]$, $\tilde{\gamma} \in (-\frac{\pi}{2}, \frac{\pi}{2})$, $\tilde{\alpha} \in [0, 2\pi)$, which has ambiguities at $\tilde{\beta} = \pm\frac{\pi}{2}$, cf. [18], see Fig. 13.

The first chart provides a diffeomorphism in an open environment around the ambiguity points of the second chart, and vice versa the second chart provides a diffeomorphism in an open environment of the first chart. A complete atlas of $SE(3)$ is thereby given by

$$(x, y, z, \alpha, \beta, \gamma) \mapsto (x, y, z, R_{\mathbf{e}_z,\gamma} R_{\mathbf{e}_y,\beta} R_{\mathbf{e}_z,\alpha}),$$
$$(x, y, z, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) \mapsto (x, y, z, R_{\mathbf{e}_x,\tilde{\gamma}} R_{\mathbf{e}_y,\tilde{\beta}} R_{\mathbf{e}_z,\tilde{\alpha}})$$

and the corresponding complete atlas of $\mathbb{R}^3 \rtimes S^2$ is given by

$$(x, y, z, \beta, \gamma) \mapsto (x, y, z, [R_{\mathbf{p}}]) \equiv (x, y, z, \cos\gamma \sin\beta, \sin\gamma \sin\beta, \cos\beta),$$
$$(x, y, z, \tilde{\beta}, \tilde{\gamma}) \mapsto (x, y, z, [\tilde{R}_{\mathbf{p}}]) \equiv (x, y, z, \sin\tilde{\beta}, -\cos\tilde{\beta}\sin\tilde{\gamma}, \cos\tilde{\beta}\cos\tilde{\gamma}).$$

The Logarithmic Map

Because $SE(3)$ is a Lie group, the exponential map is linked to the matrix exponentiation of a certain matrix representation for elements of $T_e(SE(3))$. Only the resulting expressions are given here, while the derivations can be found in Duits and Franken [18, Sect. 5.1]. The exponential map of a tangent vector $c^i(\mathbf{x}, Q)A_i$, using the short-hand notation $c^i \equiv c^i(\mathbf{x}, Q)$, is given by

$$\exp(c^i A_i) = \left(\left(I + \tfrac{1-\cos\|\mathbf{c}_2\|}{\|\mathbf{c}_2\|^2}\Omega + \tfrac{1-\operatorname{sinc}\|\mathbf{c}_2\|}{\|\mathbf{c}_2\|^2}\Omega^2\right)\cdot \mathbf{c}_1\ ,\right.$$
$$\left. I + \operatorname{sinc}(\|\mathbf{c}_2\|)\,\Omega + \tfrac{1-\cos\|\mathbf{c}_2\|}{\|\mathbf{c}_2\|^2}\Omega^2\right),$$

where $\mathbf{c}_1 \equiv \mathbf{c}_1(\mathbf{x}, Q) = (c^1, c^2, c^3)^T$ and $\mathbf{c}_2 \equiv \mathbf{c}_2(\mathbf{x}, Q) = (c^4, c^5, c^6)^T$, and with $\Omega \equiv \Omega(\mathbf{x}, Q) = \begin{pmatrix} 0 & -c^6 & c^5 \\ c^6 & 0 & -c^4 \\ -c^5 & c^4 & 0 \end{pmatrix}$. As the exponential map is invertible, the logarithmic map (the inverse of the exponential map) can be derived

$$\log(\mathbf{x}, Q) = \sum_{i=1}^{6} c^i(\mathbf{x}, Q)A_i\,.$$

This equation can be solved to give expressions for c^i in terms of the first chart, resulting in

$$\mathbf{c}_2 = \frac{1}{2\operatorname{sinc}(\tilde{q})}\begin{pmatrix} \sin\beta\,(\sin\alpha - \sin\gamma) \\ \sin\beta\,(\cos\alpha + \cos\gamma) \\ 2\cos^2\frac{\beta}{2}\sin(\alpha+\gamma) \end{pmatrix},$$

where we have $\tilde{q} = \|\mathbf{c}_2\| = \arcsin\sqrt{\cos^2\frac{\alpha+\gamma}{2}\sin^2\beta + \cos^4\frac{\beta}{2}\sin^2(\alpha+\gamma)}$, and

$$\mathbf{c}_1 = \left(I - \frac{1}{2}\Omega + \frac{1}{\tilde{q}^2}\left(1 - \frac{\tilde{q}}{2}\cot\frac{\tilde{q}}{2}\right)\Omega^2\right)\cdot\mathbf{x}.$$

This last expression is retained when rewriting the logarithmic map in terms of the second chart, but for $\tilde{q}$ and $\mathbf{c}_2$ we have

$$\tilde{q} = \arcsin\sqrt{\cos^4\tfrac{\tilde{\gamma}}{2}\sin^2\tilde{\beta} + \cos^2\tfrac{\tilde{\beta}}{2}\sin^2\tilde{\gamma}},$$

$$\mathbf{c}_2 = \tfrac{1}{\operatorname{sinc}(\tilde{q})}\begin{pmatrix} \sin\tilde{\gamma}\cos^2\frac{\tilde{\beta}}{2} \\ \sin\tilde{\beta}\cos^2\frac{\tilde{\gamma}}{2} \\ \frac{1}{2}\sin\tilde{\gamma}\sin\tilde{\beta} \end{pmatrix}.$$

Finite Difference Schemes

The finite difference stencils used in the Mathematica package are given here. Writing $Q_{\mathbf{v},\alpha}$ for a rotation by angle α around $\mathbf{v}$ expressed in the second coordinate chart, we find the explicit formulas for the vector fields on $\mathbb{R}^3 \rtimes S^2$ given below. h_s and h_a are respectively the spatial and angular step sizes. h_s is typically 0.88, see Creusen [12], while h_a depends on the distribution of sample points on the sphere. See Duits and Franken [18, Chap. 7] for the derivation of these formulas from Eq. (10).

Remark 8. The use of single values for h_s and h_a relies on approximately equidistantly sampled data.

The three different finite differences schemes considered are given below. Second order derivatives are calculated by repeating the listed first order derivatives.

Central derivatives

$$\begin{aligned}
\mathscr{A}_1 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x}+h_s Q_{\mathbf{n}}\mathbf{e}_x,\mathbf{n})-U(\mathbf{x}-h_s Q_{\mathbf{n}}\mathbf{e}_x,\mathbf{n})}{2h_s}\\
\mathscr{A}_2 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x}+h_s Q_{\mathbf{n}}\mathbf{e}_y,\mathbf{n})-U(\mathbf{x}-h_s Q_{\mathbf{n}}\mathbf{e}_y,\mathbf{n})}{2h_s}\\
\mathscr{A}_3 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x}+h_s Q_{\mathbf{n}}\mathbf{e}_z,\mathbf{n})-U(\mathbf{x}-h_s Q_{\mathbf{n}}\mathbf{e}_z,\mathbf{n})}{2h_s}\\
\mathscr{A}_4 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x},Q_{\mathbf{n}}Q_{\mathbf{e}_x,h_a}\mathbf{e}_z)-U(\mathbf{x},Q_{\mathbf{n}}Q_{\mathbf{e}_x,-h_a}\mathbf{e}_z)}{2h_a}\\
\mathscr{A}_5 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x},Q_{\mathbf{n}}Q_{\mathbf{e}_y,h_a}\mathbf{e}_z)-U(\mathbf{x},Q_{\mathbf{n}}Q_{\mathbf{e}_y,-h_a}\mathbf{e}_z)}{2h_a}
\end{aligned}$$

Forward derivatives

$$\begin{aligned}
\mathscr{A}_1 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x}+h_s Q_{\mathbf{n}}\mathbf{e}_x,\mathbf{n})-U(\mathbf{x},\mathbf{n})}{h_s}\\
\mathscr{A}_2 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x}+h_s Q_{\mathbf{n}}\mathbf{e}_y,\mathbf{n})-U(\mathbf{x},\mathbf{n})}{h_s}\\
\mathscr{A}_3 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x}+h_s Q_{\mathbf{n}}\mathbf{e}_z,\mathbf{n})-U(\mathbf{x},\mathbf{n})}{h_s}
\end{aligned}$$

Backward derivatives

$$\begin{aligned}
\mathscr{A}_1 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x},\mathbf{n})-U(\mathbf{x}-h_s Q_{\mathbf{n}}\mathbf{e}_x,\mathbf{n})}{h_s}\\
\mathscr{A}_2 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x},\mathbf{n})-U(\mathbf{x}-h_s Q_{\mathbf{n}}\mathbf{e}_y,\mathbf{n})}{h_s}\\
\mathscr{A}_3 U(\mathbf{x},\mathbf{n}) &\approx \frac{U(\mathbf{x},\mathbf{n})-U(\mathbf{x}-h_s Q_{\mathbf{n}}\mathbf{e}_z,\mathbf{n})}{h_s}
\end{aligned}$$

For the spatial derivatives, a simplified version of the upwind biased scheme is implemented in order to properly handle boundary effects. The scheme first calculates the central derivative in every point. Then for every point, depending on the sign of the central derivative, either the forward- or the backward derivative is calculated. A positive value of the central derivative means a backward derivative is calculated. The ensemble of forward- and backward derivatives in all points is then the actual derivative. The angular derivatives are all calculated using central derivatives.

Acknowledgements The authors are grateful to the Kempenhaeghe Epilepsy Center, who supplied the MRI data. Furthermore, the authors would like to thank the anonymous reviewers for their helpful comments, and the editors for their work in making this publication happen.

References

1. Aganj, I., Lenglet, C., Jahanshad, N., Yacoub, E., Harel, N., Thompson, P., Sapiro, G.: A hough transform global probabilistic approach to multiple-subject diffusion MRI tractography. Med. Image Anal. **15**(4), 414–425 (2011)
2. Alexander, D., Barker, G., Arridge, S.: Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. Magn. Reson. Med. **48**(2), 331–340 (2002)
3. Aubin, T.: A Course in Differential Geometry. American Mathematical Society, Providence (2001)
4. Basser, P., Mattiello, J., Le Bihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**, 259–267 (1994)
5. Bellaïche, B.: The tangent space in sub-Riemannian geometry. J. Math. Sci. **83**(4), 461–476 (1996)
6. Burgeth, B., Weickert, J.: An explanation for the logarithmic connection between linear and morphological systems. In: Proceedings of the 4th International Conference, Scale Space 2003, Isle of Skye, pp. 325–339. Springer (2003)
7. Burgeth, B., Breuß, M., Pizarro, L., Weickert, J.: PDE-driven adaptive morphology for matrix fields. In: Scale Space and Variational Methods in Computer Vision, Voss, pp. 247–258 (2009)
8. Burgeth, B., Didas, S., Weickert, J.: A general structure tensor concept and coherence-enhancing diffusion filtering for matrix fields. In: Laidlaw, D., Weickert, J. (eds.) Visualization and Processing of Tensor Fields: Advances and Perspectives, pp. 305–324. Springer, Berlin (2009)
9. Burgeth, B., Pizarro, L., Didas, S.: Edge-enhancing diffusion filtering for matrix fields. In: Laidlaw, D.H., Vilanova, A. (eds.) New Developments in the Visualization and Processing of Tensor Fields, pp. 51–67. Springer, Berlin/Heidelberg (2012)
10. Chirikjian, G., Kyatkin, A.: Engineering Applications of Noncommutative Harmonic Analysis: With Emphasis on Rotation and Motion Groups. CRC, Boca Raton (2000)
11. Crandall, M.G., Lions, P.L.: Viscosity solutions of Hamilton-Jacobi equations. Trans. Am. Math. Soc. **277**(1), 1–42 (1983)
12. Creusen, E.: Numerical schemes for linear and non-linear enhancement of HARDI data. Master's thesis, Eindhoven University of Tecnhology (2010)
13. Creusen, E., Duits, R., Dela Haije, T.: Numerical schemes for linear and non-linear enhancement of HARDI data. In: Proceedings of the Third International Conference on Scale Space and Variational Methods in Computer Vision, Ein-Gedi. Lecture Notes in Computer Science, vol. 12 (2011)
14. Creusen, E., Duits, R., Vilanova, A., Florack, L.: Numerical schemes for linear and non-linear enhancement of DW-MRI. Numer. Math.: Theory Methods Appl. **6**, 138–168 (2013)
15. Dela Haije, T.: Geometrical methods for enhancement and tractography in diffusion MRI. Master's thesis, Eindhoven University of Technology (2012)
16. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical Q-ball imaging. Magn. Reson. Med. **58**(3), 497–510 (2007)
17. Duits, R.: Perceptual organization in image analysis. Ph.D. thesis, Technische Universiteit Eindhoven (2005)
18. Duits, R., Franken, E.: 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**(3), 231–264 (2011)

19. Duits, R., Creusen, E., Ghosh, A., Dela Haije, T.: Diffusion, convection and erosion on $\mathbb{R}^3 \rtimes S^2$ and their application to the enhancement of crossing fibers. CASA report 18, Technische Universiteit Eindhoven (2011). http://arxiv.org/abs/1103.0656v5
20. Duits, R., Dela Haije, T., Creusen, E., Ghosh, A.: Morphological and linear scale spaces for fiber enhancement in DW-MRI (2012, to appear). J. Math. Imaging Vis. http://link.springer.com/article/10.1007%2Fs10851-012-0387-2
21. Duits, R., Dela Haije, T.C.J., Ghosh, A., Creusen, E.J., Vilanova, A., ter Haar Romeny, B.: Fiber enhancement in diffusion-weighted MRI. In: Scale Space and Variational Methods in Computer Vision. Lecture Notes in Computer Science, vol. 6667 (Springer, Berlin/Heidelberg, 2012), pp. 1–13
22. Dungey, N., ter Elst, A.F.M., Robinson, D.W.: Analysis on Lie Groups with Polynomial Growth. Progress in Mathematics, vol. 214. Birkhäuser, Boston (2003)
23. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability, vol. 57. Chapman & Hall, New York (1997)
24. Evans, L.: Partial Differential Equations. Graduate Studies in Mathematics, vol. 19. American Mathematical Society, Providence (1998)
25. Florack, L.: Codomain scale space and regularization for high angular resolution diffusion imaging. In: Aja Fernández, S., de Luis Garcia, R. (eds.) CVPR Workshop on Tensors in Image Processing and Computer Vision, Anchorage (2008)
26. Florack, L., Balmachnova-Sizykh, E.: Decomposition of high angular resolution diffusion images into a sum of self-similar polynomials on the sphere. In: Proceedings of the 18th International Conference on Computer Graphics and Vision, Moscow, pp. 26–31 (2008)
27. Franken, E.: Enhancement of crossing elongated stuctures in images. Ph.D. thesis, Technische Universiteit Eindhoven (2008)
28. Hagmann, P., Jonasson, L., Maeder, P., Thiran, J.P., Wedeen, V., Meuli, R.: Understanding diffusion MR imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond. Radiographics **26**, S205–S223 (2006)
29. Hess, C.P., Mukherjee, P., Tan, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. Magn. Reson. Med. **56**, 104–117 (2006)
30. Hörmander, L.: Hypoelliptic second order differential equations. Acta Math. **119**(1), 147–171 (1967)
31. Jackway, P.: Morphological scale-space. In: Proceedings 11th IAPR International Conference on Pattern Recognition, The Hague, pp. 252–255. Society Press (1992)
32. Jeurissen, B., Leemans, A., Jones, D., Tournier, J.D., Sijbers, J.: Probabilistic fiber tracking using the residual bootstrap with constrained spherical deconvolution. Hum. Brain Mapp. **32**, 461–479 (2011)
33. Jones, D., Cercignani, M.: Twenty-five pitfalls in the analysis of diffusion MRI data. NMR Biomed. **23**, 803–820 (2010)
34. Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M.: MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology **601**(2), 401–407 (1986)
35. Manfredi, J., Stroffolini, B.: A version of the Hopf-Lax formula in the Heisenberg group. Commun. Partial Differ. Equ. **27**(5), 1139–1159 (2002)
36. Moseley, M., Cohen, Y., Kucharchzyk, J., Mintorovitch, J., Asgari, H., Wendland, M., Tsuruda, J., Norman, D.: Diffusion-weighed MR imaging of anisotropic water diffusion in cat nervous system. Radiology **176**(2), 439–445 (1990)
37. Nagel, A., Ricci, F., Stein, E.: Fundamental solutions and harmonic analysis on nilpotent groups. Bull. Am. Math. Soc. **23**(1), 139–144 (1990)
38. Özarslan, E., Mareci, T.: Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution imaging. Magn. Reson. Med. **50**, 955–965 (2003)
39. Özarslan, E., Shepherd, T., Vemuri, B., Blackband, S., Mareci, T.: Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). NeuroImage **31**, 1086–1103 (2006)

40. Pizarro, L., Burgeth, B., Didas, S., Weickert, J.: A generic neighbourhood filtering framework for matrix fields. In: Computer Vision-ECCV 2008, Marseille, pp. 521–532. Springer (2008)
41. Prčkovska, V., Rodrigues, P., Duits, R., ter Haar Romeny, B., Vilanova, A.: Extrapolating fiber crossings from DTI data: can we gain the same information as HARDI? Technical report, Technische Universiteit Eindhoven (2010)
42. Reisert, M., Skibbe, H.: Left-invariant diffusion on the motion group in terms of the irreducible representations of SO(3). http://arxiv.org/pdf/1202.5414v1.pdf
43. Rodrigues, P., Duits, R., Vilanova, A., ter Haar Romeny, B.: Accelerated diffusion operators for enhancing DW-MRI. In: Bartz, D., Botha, C., Hornegger, J., Machiraju, R. (eds.) Eurographics Workshop on Visual Computing for Biology and Medicine, Leipzig, pp. 49–56 (2010)
44. Roerdink, J.: Group morphology. Pattern Recognit. **33**, 877–895 (2000)
45. Sherbondy, A., Doughery, R., Ben-Shachar, M., Napel, S., Wandell, B.: ConTrack: finding the most likely pathways between brain regions using diffusion tractography. J. Vis. **8**(15), 1–16 (2008)
46. Stejskal, E., Tanner, J.: Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J. Chem. Phys. **43**, 288–293 (1965)
47. Tabelow, K., Polzehl, J., Spokoiny, V., Voss, H.: Diffusion tensor imaging: structural adaptive smoothing. NeuroImage **39**(4), 1763–1773 (2008)
48. Tabelow, K., Keller, S., Mohammadi, S., Kugel, H., Gerdes, J., Polzehl, J., Deppe, M.: Structural adaptive smoothing increases sensitivity of DTI to detect microstructure alterations. In: Poster at the 17th Annual Meeting of the Organization for Human Brain Mapping, Québec City (2011)
49. Tax, C., Duits, R., Vilanova, A., ter Haar Romeny, B., Leemans, A., Ossenblok, P.: Evaluating contextual processing in diffusion MRI: application to optic radiation reconstruction for epilepsy surgery (2012). To Apear on arXiv July 2012
50. van den Boomgaard, R., Smeulders, A.: The morphological structure of images: the differential equations of morphological scale-space. IEEE Trans. Pattern Anal. Mach. Intell. **16**(11), 1101–1113 (1994)
51. Warach, S., Chien, D., Li, W., Ronthal, M., Edelman, R.: Fast magnetic resonance diffusion-weighted imaging of acute human stroke. NeuroImage **42**(9), 1717–1723 (1992)
52. Wedeen, V., Hagmann, P., Tseng, W., Reese, T., Weiskoff, R.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn. Reson. Med. **54**(6), 1377–1386 (2005)
53. Welk, M.: Families of generalised morphological scale spaces. In: Scale Space Methods in Computer Vision, Isle of Skye. Lecture Notes in Computer Science, pp. 770–784 (2003)

Part III
Higher Order Tensors and Riemannian-Finsler Geometry

Higher-Order Tensors in Diffusion Imaging

Thomas Schultz, Andrea Fuster, Aurobrata Ghosh, Rachid Deriche, Luc Florack, and Lek-Heng Lim

Abstract Diffusion imaging is a noninvasive tool for probing the microstructure of fibrous nerve and muscle tissue. Higher-order tensors provide a powerful mathematical language to model and analyze the large and complex data that is generated by its modern variants such as High Angular Resolution Diffusion Imaging (HARDI) or Diffusional Kurtosis Imaging. This survey gives a careful introduction to the foundations of higher-order tensor algebra, and explains how some concepts from linear algebra generalize to the higher-order case. From the application side, it reviews a variety of distinct higher-order tensor models that arise in the context of diffusion imaging, such as higher-order diffusion tensors, q-ball or fiber Orientation Distribution Functions (ODFs), and fourth-order covariance and kurtosis tensors. By bridging the gap between mathematical foundations and application, it provides an introduction that is suitable for practitioners and applied mathematicians alike, and propels the field by stimulating further exchange between the two.

T. Schultz (✉)
University of Bonn, Bonn, Germany

Max Planck Institute for Intelligent Systems, Tübingen, Germany
e-mail: schultz@cs.uni-bonn.de

A. Fuster (✉) • L. Florack
Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: a.fuster@tue.nl; l.m.j.florack@tue.nl

A. Ghosh (✉) • R. Deriche
INRIA Sophia Antipolis-Méditerranée, Sophia Antipolis Cedex, France
e-mail: aurobrata.ghosh@inria.fr; rachid.deriche@inria.fr

L.-H. Lim
University of Chicago, Chicago, IL, USA
e-mail: lekheng@uchicago.edu

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_6,

1 Introduction

In biological tissues such as nerve fiber bundles and muscles, the spontaneous heat motion of water molecules is restricted by obstacles in the fibrous microstructure. Diffusion Imaging [70] uses the principles of Magnetic Resonance Imaging (MRI) to non-invasively measure properties of this motion, which is also known as self-diffusion. When applied to the human brain, this provides unique insights about brain connectivity, which makes diffusion MRI one of the key technologies in an ongoing large-scale scientific effort to map the human brain connectome [33]. Consequently, it is a timely and important topic of research to create mathematical models that infer biologically meaningful parameters from such data.

Higher-order tensors have been used in applications ranging from psychometrics [64] and chemometrics [103] to signal processing [102], computer vision [110], and neuroscience [85]. They also provide adequate models for a number of quantities that occur in the context of diffusion imaging. Many practitioners view higher-order tensors as a generalization of matrices to multi-way arrays. However, tensors can also be studied in an invariant, coordinate-free notation. Tensor decompositions are an active and challenging topic in applied mathematics, since fundamental concepts from linear algebra, such as the singular value decomposition, do not have unique generalization to higher order, and most generalizations are hard to compute.

It is a goal of our survey to stimulate an active exchange between mathematicians, who are studying tensor decompositions and the geometry of tensors, and computer scientists and MR physicists, who are interested in using tensors as mathematical tools in the context of diffusion MRI. Therefore, unlike previous surveys [39, 90], Sect. 2 provides a broad overview of all physical quantities that have been modeled with higher-order tensors in the context of diffusion MRI. On the other hand, our introduction to the higher-order tensor formalism in Sect. 3 differs from existing discussions [66, 73] by focusing on aspects relevant to this specific application.

Relevant literature is spread over journals in applied mathematics, MR physics, neuroimaging, and computer science. Drawing on all these fields, Sect. 4 presents the current state of the art on fitting higher-order tensor models to the measured data, and Sect. 5 discusses operations performed on the tensors for further analysis. Among others, this includes computation of scalar invariants (Sect. 5.1), maximum detection (Sect. 5.3), and tensor decompositions (Sect. 5.4).

2 Overview of Higher-Order Tensor Models in dMRI

Different physical quantities that can be measured by or inferred from diffusion MRI have been modeled with higher-order tensors. The resulting tensors not only differ in their interpretation, but also in dimension, order, and symmetry.

Diffusion imaging inserts magnetic field gradients into the MR sequence which sensitize the measurement to molecular motion along the gradient direction [70].

Compared to an image without diffusion weighting, this leads to an attenuation of signal strength. The standard diffusion tensor model [18] assumes that the diffusion-weighted MR signal in direction $\mathbf{u}$ is given by a monoexponential attenuation of the unweighted signal S_0, depending on the diffusion weighting b and a directionally dependent apparent diffusion coefficient, modeled by a diffusion tensor $\mathbf{D}$:

$$S(\mathbf{u}) = S_0 \mathrm{e}^{-b\mathbf{u}^{\mathrm{T}}\mathbf{D}\mathbf{u}} \tag{1}$$

Estimating the six unique coefficients of $\mathbf{D}$ requires measurements in at least six different gradient directions. Typical parameter values are $b \in [700, 1{,}000]\,\mathrm{mm}^2/\mathrm{s}$, and a spatial resolution of around $2{\times}2{\times}2\,\mathrm{mm}^3$. When studying the human brain, this corresponds to a subdivision into around 10^5 volume elements (voxels); a separate diffusion tensor $\mathbf{D}$ is computed for each of them.

Since nerve fibers are on the micrometer scale and therefore far below image resolution, their complex organization often leads to apparent diffusivities $D(\mathbf{u})$ that are poorly approximated by a quadratic function. For these cases, Eq. (1) has been generalized to use higher-order polynomials. As it will be explained in Sect. 3.3, this corresponds to a higher-order diffusion tensor $\mathscr{D}$ [86]:

$$S(\mathbf{u}) = S_0 \mathrm{e}^{-b\,D(\mathbf{u})} \quad \text{with} \quad D(\mathbf{u}) = \mathscr{D} \cdot^k \mathbf{u} \tag{2}$$

Such High Angular Resolution Diffusion Imaging (HARDI) models require a larger number of 30–100 gradient directions, and larger $b \in [1{,}000, 3{,}000]\,\mathrm{mm}^2/\mathrm{s}$.

One goal in diffusion imaging is to estimate the dominant nerve fiber directions within each voxel. When there is only one such direction, the principal eigenvector of the diffusion tensor $\mathbf{D}$ is aligned with it. However, a mixture of multiple fiber directions is not easily resolved with the higher-order diffusion tensor $\mathscr{D}$. For this purpose, it is easier to consider the diffusion propagator $P(\mathbf{x})$, the probability density of a molecular displacement along vector $\mathbf{x}$ within the diffusion time. Under certain assumptions, $P(\mathbf{x})$ can be computed from $\mathscr{D}$; this will be the topic of Sect. 5.2.

Writing the diffusion propagator $P(\mathbf{x})$ in spherical coordinates and integrating over the radius results in the diffusion orientation distribution function $\psi(\mathbf{u})$, whose maxima approximate the main nerve fiber directions. The q-ball model has been introduced as an approximative way of computing $\psi(\mathbf{u})$ [109]. Even though its exact interpretation has been disputed [17], q-ball maxima indicate approximate fiber directions, and q-balls are sometimes expressed in a tensor basis [7,55], making it relevant to compute the maxima of homogeneous forms (cf. Sect. 5.3).

When measuring at different b values, it is common to observe that the true signal attenuation is not monoexponential, as assumed by Eqs. (1) and (2). This indicates that the diffusion propagator $P(\mathbf{x})$ is non-Gaussian. Accounting for all higher-order moments of P leads to a different generalization of Eq. (1) [77, 78],

$$S(\mathscr{B}) = S_0\, \mathrm{e}^{\sum_{k=2}^{\infty} j^k \langle \mathscr{D}^{(k)}, \mathscr{B}^{(k)} \rangle}, \tag{3}$$

where j is the imaginary unit, $\mathscr{D}^{(k)}$ is a series of diffusion tensors with increasing order k, and the diffusion-weighted signal $S(\mathscr{B})$ is a function of a series of tensors $\mathscr{B}^{(k)}$, which combine information about the direction and strength of the diffusion weighting. $\langle \mathscr{D}^{(k)}, \mathscr{B}^{(k)} \rangle$ denotes the scalar product of the two tensors.

In contrast to Eq. (2), which uses a single higher-order tensor $\mathscr{D}$ that contains all the information that would be present in lower-order approximations, each element in the series of tensors $\mathscr{D}^{(k)}$ in Eq. (3) contains non-redundant information that is independent from all other orders k. This additional information needs to be acquired by sampling multiple b values in several gradient directions [79].

The tensors in Eq. (3) are three-dimensional, and symmetric under all index permutations. The odd orders k in Eq. (3) carry information about asymmetries in the diffusion propagator, i.e., $P(-\mathbf{x}) \neq P(\mathbf{x})$. However, that information resides in the phase of the complex-valued MR signal. At the technical state of the art, signal phase in diffusion MRI is so heavily corrupted by measurement noise and artifacts that it is not informative. Therefore, practical implementations of this generalization are limited to estimating even-order tensors from the signal magnitude [80].

Diffusional Kurtosis Imaging augments the second-order diffusion tensor $\mathbf{D}$ in Eq. (1) with a fourth-order kurtosis tensor $\mathscr{W}$ [61],

$$S(\mathbf{u}, b) = S_0 \mathrm{e}^{-b\mathbf{u}^{\mathrm{T}}\mathbf{D}\mathbf{u} + \frac{1}{6}b^2\left(\frac{1}{3}\mathrm{tr}(\mathbf{D})\right)^2 \mathscr{W} \cdot^4 \mathbf{u}}, \tag{4}$$

where tr indicates matrix trace. Computing the parameters in Eq. (4) requires measurements at multiple b values, but no signal phase. They capture the same information present in the second and fourth moments of $P(\mathbf{x})$, but allow for simpler computation of the apparent diffusional kurtosis K_{app} in direction $\mathbf{u}$:

$$K_{\mathrm{app}}(\mathbf{u}) = \frac{\left(\frac{1}{3}\mathrm{tr}(\mathbf{D})\right)^2}{\left(\mathbf{u}^{\mathrm{T}}\mathbf{D}\mathbf{u}\right)^2} \mathscr{W} \cdot^4 \mathbf{u} \tag{5}$$

For Gaussian diffusion, $K_{\mathrm{app}} = 0$. Negative kurtosis is expected from diffusion restricted by spherical pores, and positive kurtosis can indicate presence of heterogeneous diffusion compartments [61].

Fourth-order covariance tensors Σ occur in statistical models of second-order diffusion tensors [19]. Even though they are three-dimensional in each mode, they only possess partial symmetries ($\Sigma_{ijkl} = \Sigma_{klij}$; $\Sigma_{ijkl} = \Sigma_{jikl}$; $\Sigma_{ijkl} = \Sigma_{ijlk}$) [20].

If we assume that all nerve fiber bundles within a voxel have approximately the same diffusion characteristics, the MR signal is given by the convolution of a fiber orientation density function (fODF) with a kernel describing the single fiber response [107]. Unlike the diffusion ODF, values of the fODF $F(\mathbf{u})$ are interpreted as the fraction of fibers aligned with direction $\mathbf{u}$. $F(\mathbf{u})$ can be obtained by spherical deconvolution and a variant of that technique, which will be explained in Sect. 4.3, allows for further analysis of the fODF via tensor decomposition [100].

3 Mathematical Background

We include a basic introduction to tensors and tensor fields. In a nutshell, a tensor of order p or p-tensor is a multilinear functional on p vector spaces $T : \mathbb{V} \times \mathbb{V} \times \cdots \times \mathbb{V} \to \mathbb{R}$, and can be represented in coordinates as a p-dimensional matrix $A \in \mathbb{R}^{n \times n \times \cdots \times n}$, $n = \dim(\mathbb{V})$, if one chooses a basis on $\mathbb{V}$. A tensor field is a tensor-valued function on a manifold. We refer readers who are interested in further properties of tensors and hypermatrices to [73] for an elementary treatment. Mathematically sophisticated readers may consult [66] for a much more in-depth treatment.

3.1 Basic Definitions

Let us first define our basic mathematical objects: (i) tensors, and (ii) tensor fields. Let $\mathbb{V}$ be a vector space over $\mathbb{R}$. An order-p *tensor* is a *multilinear functional*

$$f : \underbrace{\mathbb{V} \times \mathbb{V} \times \cdots \times \mathbb{V}}_{p \text{ times}} \to \mathbb{R}.$$

Multilinear means that if all arguments are kept constant but one, then f is linear in that varying argument, i.e.,

$$f(\mathbf{u}_1, \ldots, \alpha \mathbf{v}_i + \beta \mathbf{w}_i, \ldots, \mathbf{u}_p) = \alpha f(\mathbf{u}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{u}_p) + \beta f(\mathbf{u}_1, \ldots, \mathbf{w}_i, \ldots, \mathbf{u}_p), \tag{6}$$

for every $i = 1, \ldots, p$, $\alpha, \beta \in \mathbb{R}$ and $\mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i \in \mathbb{V}$. The set of all p-tensors is called the p-fold *tensor product* of the vector space $\mathbb{V}$ and denoted

$$\mathbb{V}^{\otimes p} = \underbrace{\mathbb{V} \otimes \mathbb{V} \otimes \cdots \otimes \mathbb{V}}_{p \text{ times}}.$$

We ignore the distinction between covariant, contravariant and mixed tensors, since it is less relevant when working with coordinate representations in an orthonormal basis, as will be the case in this survey. An abstract approach towards tensors is now standard in any basic graduate courses in algebra [57, 68] or even mathematical methods courses for physicists [38]. However, such courses focus almost exclusively on properties of an entire space of tensors [116] as opposed to properties of an individual tensor, i.e., a specific element from such a tensor space. Properties of an individual tensor such as rank, norm, eigenvalues, decompositions, are of great relevance to us and will be discussed after we introduce tensor fields.

We will be informal in our treatment of tensor fields to make it more easily accessible. Readers who wish to see a rigorous definition would have no shortage of standard Refs. [24, 67, 112] to consult. Let M be a topological manifold which

we may later endow with additional structures (differential, Riemannian, Finsler, etc.). A tensor field is, roughly speaking, a tensor-valued function $F : M \to \mathbb{V}^{\otimes p}$ or alternatively, a function of the form

$$F : M \times \underbrace{\mathbb{V} \times \mathbb{V} \times \cdots \times \mathbb{V}}_{p \text{ times}} \to \mathbb{R} \tag{7}$$

with the property that for every point $\mathbf{x} \in M$,

$$F(\mathbf{x}; \cdot, \cdot, \ldots, \cdot) : \mathbb{V} \times \mathbb{V} \times \cdots \times \mathbb{V} \to \mathbb{R}$$

is a multilinear functional, i.e., $F(\mathbf{x}; \mathbf{u}_1, \ldots, \mathbf{u}_p)$ is multilinear in the last p arguments for every fixed $\mathbf{x} \in M$. If we want F to have additional properties like continuity or differentiability, this definition is only good locally, i.e., every $\mathbf{x}_0 \in M$ has a neighborhood $U_{\mathbf{x}_0} \subseteq M$ such that

$$F : U_{\mathbf{x}_0} \times \mathbb{V} \times \mathbb{V} \times \cdots \times \mathbb{V} \to \mathbb{R}$$

is multilinear for every $\mathbf{x} \in U_{\mathbf{x}_0}$. By far the most common choice for $\mathbb{V}$ is $T_{\mathbf{x}}(M)$, the tangent space at $\mathbf{x}$, i.e., the vector space $\mathbb{V}_i$ changes with each $\mathbf{x}$ and we really have a multilinear function

$$F(\mathbf{x}; \cdot, \cdot, \ldots, \cdot) : T_{\mathbf{x}}(M) \times \cdots \times T_{\mathbf{x}}(M) \to \mathbb{R}$$

at each $\mathbf{x} \in U_{\mathbf{x}_0}$. So each $F(\mathbf{x}; \cdot, \cdot, \ldots, \cdot)$ has a different domain, and F is really a family of multilinear functionals parameterized by $\mathbf{x} \in M$. The proper treatment is to define F as a *section* (of a tensor product of vector bundles) as opposed to a function (with values in a tensor product of vector spaces). In fact, tensor fields are more than pointwise multilinear functionals, they satisfy the multilinearity condition in Eq. (6) with coefficients α, β being real-valued functions on M (usually in $C^\infty(M)$ if M is a smooth manifold) instead of merely being constants in $\mathbb{R}$.

The above discussions use the coordinate-free language of modern treatments of tensors and tensor fields in mathematics. In applications such as those considered in this survey, computations require introducing coordinates by chosing a basis on $\mathbb{V}$. If we pick a basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$, where $n = \dim(\mathbb{V})$, then a multilinear functional f may be represented as an $n \times n \times \cdots \times n$ (p times) array of elements of $\mathbb{R}$:

$$\mathscr{A} = (a_{i_1 i_2 \cdots i_p})_{i_1, \ldots, i_p = 1}^{n} \in \mathbb{R}^{n \times \cdots \times n}. \tag{8}$$

We shall use the term *hypermatrix* of order p, or simply p-hypermatrix, when referring to a p-dimensional matrix of the form in Eq. (8). The origin of this terminology would appear to be [37]. These objects are natural multilinear generalizations of matrices in the following way. Since we have fixed a basis, every vector in $\mathbb{V}$ has a coordinate representation and we may assume that $\mathbb{V} = \mathbb{R}^n$.

A bilinear functional $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ can be encoded by a matrix $\mathbf{A} = [a_{ij}]_{i,j=1}^n \in \mathbb{R}^{n\times n}$, in which the entry a_{ij} records the value of $f(\mathbf{e}_i, \mathbf{e}_j) \in \mathbb{R}$ where $\mathbf{e}_i$ denotes the ith standard basis vector in $\mathbb{R}^n$. By linearity in each coordinate, specifying $\mathbf{A}$ determines the values of f on all of $\mathbb{R}^n \times \mathbb{R}^n$; in fact, we have $f(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T\mathbf{A}\mathbf{v}$ for any (column) vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Thus, matrices encode all bilinear functionals. If $\mathbf{A} = \mathbf{A}^T$ is symmetric, the corresponding bilinear map is invariant under exchanging of coordinates:

$$f(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T\mathbf{A}\mathbf{v} = (\mathbf{u}^T\mathbf{A}\mathbf{v})^T = \mathbf{v}^T\mathbf{A}^T\mathbf{u} = \mathbf{v}^T\mathbf{A}\mathbf{u} = f(\mathbf{v}, \mathbf{u}).$$

To avoid sub-subscripts, we will restrict our discussion to 4-tensors. A 4-tensor is a quadrilinear functional $f : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ which has a coordinate representation given by a 4-hypermatrix $\mathscr{A} = (a_{ijkl})_{i,j,k,l=1}^n \in \mathbb{R}^{n\times n\times n\times n}$ as in Eq. (8) with $p = 4$. The subscripts and superscripts in Eq. (8) will be dropped whenever the range of i, j, k, l is obvious or unimportant. A 4-hypermatrix is said to be *symmetric* if the value of a_{ijkl} stays the same for all 24 permutations of the indices:

$$a_{ijkl} = a_{ijlk} = a_{jilk} = \cdots = a_{lkji}.$$

Symmetric 4-tensors correspond to coordinate representations of quadrilinear maps $f : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ with

$$f(\mathbf{t}, \mathbf{u}, \mathbf{v}, \mathbf{w}) = f(\mathbf{t}, \mathbf{u}, \mathbf{w}, \mathbf{v}) = f(\mathbf{u}, \mathbf{t}, \mathbf{v}, \mathbf{w}) = \cdots = f(\mathbf{w}, \mathbf{v}, \mathbf{u}, \mathbf{t}).$$

The set of symmetric 4-hypermatrices is often denoted $\mathsf{S}^4(\mathbb{R}^n)$ and it forms a linear subspace of the vector space $\mathbb{R}^{n\times n\times n\times n}$. More generally $\mathsf{S}^p(\mathbb{V})$, the set of symmetric p-tensors over an arbitrary vector space $\mathbb{V}$, may be defined in a coordinate-free manner [26] and forms a subspace of $\mathbb{V}^{\otimes p}$.

What about tensor fields? Since any manifold M may be given local coordinates, we may view tensor fields as hypermatrix-valued functions $F : M \to \mathbb{R}^{n\times\cdots\times n}$, $\mathbf{x} \mapsto \mathscr{A}_\mathbf{x} = (a_{i_1 i_2 \cdots i_p}(\mathbf{x}))_{i_1,\ldots,i_p=1}^n$, that are locally defined (roughly speaking, they are defined for local coordinates chosen for each neighborhood $U_\mathbf{x} \subseteq M$). The coordinate-dependent view of tensor fields as (hyper)matrix-valued functions is the classical approach. The subject, studied in this light, is often called tensor calculus, tensor analysis, or Ricci calculus. Tullio Levi-Civita, Gregorio Ricci-Curbastro, and Jan Schouten are usually credited for its invention [104].

3.2 Tensor Algebra and Homogeneous Polynomials

As we saw in the last section, a 4-hypermatrix $\mathscr{A} \in \mathbb{R}^{n\times n\times n\times n}$, is a coordinate representation of a 4-tensor, i.e., a quadrilinear functional $f : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. The set of 4-hypermatrices is naturally equipped with algebraic operations

inherited from the algebraic structure of the tensor product space $\mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$:

- *Addition and Scalar Multiplication:* for $(a_{ijkl}), (b_{ijkl}) \in \mathbb{R}^{n\times n\times n\times n}$ and $\lambda, \mu \in \mathbb{R}$,

$$\lambda(a_{ijkl}) + \mu(b_{ijkl}) = (\lambda a_{ijkl} + \mu b_{ijkl}) \in \mathbb{R}^{n\times n\times n\times n}, \tag{9}$$

- *Outer Product Decomposition:* every $\mathscr{A} = (a_{ijkl}) \in \mathbb{R}^{n\times n\times n\times n}$ may be decomposed as

$$\mathscr{A} = \sum\nolimits_{q=1}^{r} \lambda_q \, \mathbf{w}_q \otimes \mathbf{x}_q \otimes \mathbf{y}_q \otimes \mathbf{z}_q, \qquad a_{ijkl} = \sum\nolimits_{q=1}^{r} \lambda_q w_{iq} x_{jq} y_{kq} z_{lq}, \tag{10}$$

  with $\lambda_q \in \mathbb{R}$, $\mathbf{w}_q, \mathbf{x}_q, \mathbf{y}_q, \mathbf{z}_q \in \mathbb{R}^n$ for $q = 1, \ldots, r$. The symbol $\otimes$ here denotes the *Segre outer product*: for vectors $\mathbf{w} = [w_1, \ldots, w_n]^T, \ldots, \mathbf{z} = [z_1, \ldots, z_n]^T$,

$$\mathbf{w} \otimes \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z} := (w_i x_j y_k z_l)_{i,j,k,l=1}^{n} \in \mathbb{R}^{n\times n\times n\times n},$$

  with obvious generalization to an arbitrary number of vectors. The ℓ-fold outer product of $\mathbf{x}$ with itself is written $\mathbf{x}^{\ell}$.
- *Multilinear Matrix Multiplication:* every $\mathscr{A} = (a_{ijkl}) \in \mathbb{R}^{n\times n\times n\times n}$ may be multiplied on its '4 sides' by matrices $\mathbf{W} = [w_{i\alpha}]$, $\mathbf{X} = [x_{j\beta}]$, $\mathbf{Y} = [y_{k\gamma}]$, $\mathbf{Z} = [z_{l\delta}] \in \mathbb{R}^{n\times r}$ as follows

$$\mathscr{A} \cdot (\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = (c_{\alpha\beta\gamma\delta})_{\alpha,\beta,\gamma,\delta=1}^{n} \in \mathbb{R}^{n\times n\times n\times n}, \tag{11}$$

$$c_{\alpha\beta\gamma\delta} = \sum\nolimits_{i,j,k,l=1}^{n} a_{ijkl} w_{i\alpha} x_{j\beta} y_{k\gamma} z_{l\delta}.$$

A different choice of bases $\mathbf{b}'_1, \ldots, \mathbf{b}'_n$ on $\mathbb{V}$ would lead to a different hypermatrix representation $\mathscr{B} \in \mathbb{R}^{n\times n\times n\times n}$ of elements in $\mathbb{V} \otimes \mathbb{V} \otimes \mathbb{V} \otimes \mathbb{V}$ – where the two hypermatrix representations $\mathscr{A}$ and $\mathscr{B}$ would be related precisely by a multilinear matrix multiplication of the form

$$\mathscr{A} \cdot (\mathbf{X}, \mathbf{X}, \mathbf{X}, \mathbf{X}) = \mathscr{B}$$

where $\mathbf{X}$ is the change-of-basis matrix, i.e., an invertible matrix with $\mathbf{X}\mathbf{b}_q = \mathbf{b}'_q$ for $q = 1, \ldots, n$. Therefore, a tensor and a hypermatrix are different in the same way a linear operator and a matrix are different. Note that in the context of matrices,

$$\mathbf{x} \otimes \mathbf{y} = \mathbf{x}\mathbf{y}^T \quad \text{and} \quad \mathbf{A} \cdot (\mathbf{X}, \mathbf{Y}) = \mathbf{Y}^T \mathbf{A} \mathbf{X}.$$

When $r = 1$ in Eq. (11), i.e., the matrices $\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}$ are vectors $\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}$, we omit the $\cdot$ and write

$$\mathscr{A}(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum\nolimits_{i,j,k,l=1}^{n} a_{ijkl} w_i x_j y_k z_l \tag{12}$$

for the associated quadrilinear functional. Another special case occurs when one or more of the matrices $\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}$ in Eq. (11) is the identity $\mathbf{I} = \mathbf{I}_{n\times n}$. For example,

$$\mathscr{A}(\mathbf{I}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum\nolimits_{j,k,l=1}^{n} a_{ijkl} x_j y_k z_l \in \mathbb{R}^n. \tag{13}$$

In particular, the (partial) gradient of the quadrilinear functional $\mathscr{A}(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z})$ may be expressed as

$$\nabla_{\mathbf{w}} \mathscr{A}(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathscr{A}(\mathbf{I}, \mathbf{x}, \mathbf{y}, \mathbf{z}), \quad \nabla_{\mathbf{x}} \mathscr{A}(\mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathscr{A}(\mathbf{w}, \mathbf{I}, \mathbf{y}, \mathbf{z}), \quad \text{etc.}$$

For a symmetric 4-tensor $\mathscr{S}$, we write $\mathscr{S} \cdot \mathbf{x}$ as a shorthand for $\mathscr{S}(\mathbf{x}, \mathbf{I}, \mathbf{I}, \mathbf{I})$; the result is a 3-tensor. Repeating this operation ℓ times is written $\mathscr{S} \cdot^{\ell} \mathbf{x}$. With this notation, the homogeneous quartic polynomial $\mathscr{S}(\mathbf{x})$ that is uniquely associated with $\mathscr{S}$ can be written as

$$\mathscr{S}(\mathbf{x}) := \mathscr{S}(\mathbf{x}, \mathbf{x}, \mathbf{x}, \mathbf{x}) = \mathscr{S} \cdot^4 \mathbf{x} = \sum\nolimits_{d_1+\cdots+d_n=4} \mu_{d_1\cdots d_n} \sigma_{d_1\cdots d_n} x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n}. \tag{14}$$

Similarly, the gradient of $\mathscr{S}(\mathbf{x})$ can be conveniently expressed as $\nabla \mathscr{S}(\mathbf{x}) = 4\mathscr{S} \cdot^3 \mathbf{x}$. The right-hand side of Eq. (14) is the more typical way of writing a homogeneous polynomial in terms of monomials, unique coefficients $\sigma_{d_1,\ldots,d_n}$, and multiplicities $\mu_{d_1,\ldots,d_n} := \binom{n}{d_1,\ldots,d_n}$. This is the higher-order equivalent of writing, for $\mathbf{A} = \left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]$ and $\mathbf{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$,

$$\mathbf{A}(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = a x_1^2 + b x_1 x_2 + b x_2 x_1 + c x_2^2 = a x_1^2 + 2b x_1 x_2 + c x_2^2.$$

The *Frobenius norm* or *Hilbert-Schmidt norm* of a tensor $\mathscr{A}$ is defined by

$$\|\mathscr{A}\|_F^2 = \sum\nolimits_{i,j,k,l=1}^{n} a_{ijkl}^2. \tag{15}$$

This is by far the most popular choice of norms used for a tensor since it is readily computable and also because it is induced by an inner product

$$\langle \mathscr{A}, \mathscr{B} \rangle = \sum\nolimits_{i,j,k,l=1}^{n} a_{ijkl} b_{ijkl} \tag{16}$$

that generalizes the trace inner product. For symmetric p-tensors $\mathscr{S}, \mathscr{T}$ expressed in monomial form as in Eq. (14), this inner product may be written in the form

$$\langle \mathscr{S}, \mathscr{T} \rangle := \sum\nolimits_{d_1+\cdots+d_n=p} \mu_{d_1\cdots d_n} \sigma_{d_1\cdots d_n} \tau_{d_1\cdots d_n}$$

and is often called the *apolar inner product* in invariant theory. For any $\mathbf{v} \in \mathbb{R}^n$, the apolar inner product of a symmetric tensor and the rank-1 symmetric tensor $\mathbf{v}^{\otimes p} := \mathbf{v} \otimes \cdots \otimes \mathbf{v}$ (p times),

$$\langle \mathscr{S}, \mathbf{v}^{\otimes p} \rangle = \mathscr{S}(\mathbf{v}),$$

which makes the set of symmetric p-tensors into a reproducing kernel Hilbert space.

3.3 Homogeneous Polynomials and Spherical Harmonics

By restricting Eq. (14) to the 3D unit sphere S^2, $\mathbf{x} = [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]^T$, every symmetric tensor $\mathscr{S}$ defines a real-valued homogeneous polynomial function on S^2. Spherical harmonics (SH) are an alternate basis for describing functions on the sphere. The SHs form a complex complete orthonormal basis for square integrable functions on the unit sphere. Spherical functions can, therefore, be naturally expanded in the infinite SH basis or approximated to any accuracy by a truncated series. Again the diffusion signal being real and symmetric, a modified real and symmetric SH basis is chosen in dMRI. Therefore, S can be written as

$$S(\theta,\phi) = \sum\nolimits_{j=1}^{M'} c_j Y_j(\theta,\phi), \tag{17}$$

where $\theta \in [0,\pi]$, $\phi \in [0,2\pi)$ and c_j are the coefficients describing S in the modified SH basis [29]

$$Y_j(\theta,\phi) = \begin{cases} \sqrt{2}\,\mathrm{Re}(Y_l^{|m|}(\theta,\phi)) & \text{if } m < 0, \\ Y_l^m(\theta,\phi) & \text{if } m = 0, \\ (-1)^{m+1}\sqrt{2}\,\mathrm{Im}(Y_l^m(\theta,\phi)) & \text{if } m > 0, \end{cases} \tag{18}$$

with $Y_l^m(\theta,\phi)$ the rank l and degree m regular complex spherical harmonic:

$$Y_l^m(\theta,\phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta) e^{im\phi}, \quad m \le |l|. \tag{19}$$

In [28, 86] it was shown that the tensor basis and the SH basis are bijective via a linear transformation when the rank l of the truncated SH basis equals the order k of the symmetric tensor. This can be understood from the spherical harmonic transform of the polynomial representation of S:

$$c_j = \sum\nolimits_{i=1}^{M'=M} \mu_i \sigma_i \int_{S_2} x_1^{\alpha_i} x_2^{\beta_i} x_3^{l-\alpha_i-\beta_i} Y_j(\theta,\phi) d\Omega. \tag{20}$$

where the new indexing of μ and σ assumes an arbitrary ordering of the $\mu_{d_1 \cdots d_n}$ and $\sigma_{d_1 \cdots d_n}$ from Eq. (14). Since the integral does not depend on the tensor coefficients σ_j, Eq. (20) can be seen as a dot product between the vector of unique tensor coefficients and the vector of spherical harmonic transforms of the M monomials $x_1^{\alpha_i} x_2^{\beta_i} x_3^{l-\alpha_i-\beta_i}$. In other words, computing the M SH coefficients can be written as a matrix vector multiplication

$$\mathbf{c} = \mathbf{M}\mathbf{s}, \tag{21}$$

where $\mathbf{c} = [c_1, c_2, \ldots, c_M]^T$, $\mathbf{s} = [\sigma_1, \sigma_2, \ldots, \sigma_M]^T$, and:

$$\mathbf{M} = \begin{bmatrix} \mu_1 \int_{S_2} x_1^{\alpha_1} x_2^{\beta_1} x_3^{l-\alpha_1-\beta_1} Y_1 d\Omega & \ldots & \mu_M \int_{S_2} x_1^{\alpha_M} x_2^{\beta_M} x_3^{l-\alpha_M-\beta_M} Y_1 d\Omega \\ \vdots & \ddots & \vdots \\ \mu_1 \int_{S_2} x_1^{\alpha_1} x_2^{\beta_1} x_3^{l-\alpha_1-\beta_1} Y_M d\Omega & \ldots & \mu_M \int_{S_2} x_1^{\alpha_M} x_2^{\beta_M} x_3^{l-\alpha_M-\beta_M} Y_M d\Omega \end{bmatrix}. \tag{22}$$

3.4 Tensor Decompositions and Approximations

A tensor that can be expressed as an outer product of vectors is called *decomposable* and rank-1 if it is also nonzero. More generally, the *rank* of a tensor $\mathscr{A} = (a_{ijkl})_{i,j,k,l=1}^{n} \in \mathbb{R}^{n \times n \times n \times n}$, denoted $\text{rank}(\mathscr{A})$, is defined as the minimum r for which $\mathscr{A}$ may be expressed as a sum of r rank-1 tensors [52, 53],

$$\text{rank}(\mathscr{A}) := \min \left\{ r \;\middle|\; \mathscr{A} = \sum\nolimits_{q=1}^{r} \lambda_q \, \mathbf{w}_q \otimes \mathbf{x}_q \otimes \mathbf{y}_q \otimes \mathbf{z}_q \right\} \tag{23}$$

where the minimum is taken over all decomposition with $\lambda_p \in \mathbb{R}$, $\mathbf{w}_p, \mathbf{x}_p, \mathbf{y}_p, \mathbf{z}_p \in \mathbb{R}^n$, $p = 1, \ldots, r$. If $\mathscr{S}$ is a symmetric tensor, then its *symmetric rank* [26] is

$$\text{srank}(\mathscr{S}) := \min \left\{ r \;\middle|\; \mathscr{S} = \sum\nolimits_{q=1}^{r} \lambda_q \, \mathbf{x}_q \otimes \mathbf{x}_q \otimes \mathbf{x}_q \otimes \mathbf{x}_q \right\}. \tag{24}$$

We remark that it is not known whether the rank of a symmetric tensor is equal to its symmetric rank. The definition of rank in Eq. (23) agrees with matrix rank when applied to an order-2 tensor. In certain other literature, for example [90], the term 'rank' is used synonymously with what we called 'order' in the first paragraph of this section. For tensors of order greater than 2, rank becomes a more intricate notion than matrix rank with properties that may seem surprising at first encounter. We refer the readers to [31] (rank) and [26] (symmetric rank) for further information.

Best rank-r approximations

$$\text{argmin}_{\boldsymbol{\lambda} \in \mathbb{R}^r,\, \mathbf{W},\mathbf{X},\mathbf{Y},\mathbf{Z} \in \mathbb{R}^{n \times r}} \left\| \mathscr{A} - \sum\nolimits_{q=1}^{r} \lambda_q \, \mathbf{w}_q \otimes \mathbf{x}_q \otimes \mathbf{y}_q \otimes \mathbf{z}_q \right\| \tag{25}$$

and the corresponding best symmetric rank-r approximation problem (i.e., when $\mathbf{W} = \mathbf{X} = \mathbf{Y} = \mathbf{Z}$) are used in practice (Sect. 5.4), but have no solution in general when $r > 1$. The easiest way to explain this is that the *infimum* of the objective function, taken over all $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r) \in \mathbb{R}^r$ and $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_r]$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_r]$, $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_r]$, $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_r] \in \mathbb{R}^{n \times r}$ need not be attained. This happens regardless of symmetry, the choice of norms in Eq. (25) and for any order $p \geq 3$. In the unsymmetric case, it is known that the set of tensors of rank $s > r$ that do not have a best rank-r approximation could form a positive volume set. A particularly egregious case is $\mathbb{R}^{2\times 2\times 2}$, where no rank-3 tensor has a best rank-2 approximation. Fortunately, there are special cases where the problem can be alleviated, notably: (i) when all coordinates of $\mathscr{A}$ are nonnegative and $\boldsymbol{\lambda}, \mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z} \geq 0$ [74]; (ii) when $\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}$ satisfy a 'coherence' condition [75]; (iii) when p is even and $\boldsymbol{\lambda} \geq 0$ [76]. Unlike cases (i) and (ii), case (iii) only applies to symmetric approximations.

3.5 Eigenvectors and Eigenvalues

The basic notions for eigenvalues of tensors were introduced independently by Lim [72] and Qi [92]. The usual eigenvalues and eigenvectors of a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ are the stationary values and stationary points of its Rayleigh quotient, and this point of view generalizes naturally to tensors of higher order. This gives, for example, an *eigenvector* of a tensor $\mathscr{A} = (a_{ijkl})_{i,j,k,l=1}^{n} \in \mathbb{R}^{n\times n\times n\times n}$ as a nonzero column vector $\mathbf{x} = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$ satisfying

$$\sum\nolimits_{i,j,k=1}^{n} a_{ijkl} x_i x_j x_k = \lambda x_l, \quad l = 1, \ldots, n, \tag{26}$$

for some $\lambda \in \mathbb{R}$, which is called an *eigenvalue* of $\mathscr{A}$. Notice that if $(\lambda, \mathbf{x})$ is an eigenpair, then so is $(t^2\lambda, t\mathbf{x})$ for any $t \neq 0$; thus, eigenpairs are more naturally defined projectively. As in the matrix case, generic tensors over $\mathbb{R}$ or $\mathbb{C}$ have a finite number of eigenvalues and eigenvectors (up to this scaling equivalence), although their count is exponential in n. Still, it is possible for a tensor to have an infinite number of eigenvalues, but in that case they comprise a cofinite set of complex numbers. For an even-ordered symmetric tensor $\mathscr{S} \in \mathsf{S}^{2p}(\mathbb{R}^n)$, one has that $\mathscr{S}$ is nonnegative definite, i.e., $\mathscr{S}(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$, if and only if all the eigenvalues of $\mathscr{S}$ are nonnegative [92] – a generalization of a well-known fact for symmetric tensors.

It is worth noting that unlike in the matrix case, most tensor problems are NP-hard. This includes determining rank, best rank-1 approximation, spectral norm, eigenvalues, and eigenvectors [51]. However, the notion of NP-hardness is an asymptotic one that applies when $n \to \infty$. Therefore, these hardness results do not preclude the existence of efficient algorithms for a fixed n, and especially for small values such as $n = 3$, the case of greatest interest to diffusion MRI.

4 Fitting Higher-Order Tensor Models

4.1 Fitting Models of Apparent Diffusivity

One of the earliest models that attempted to overcome the limitations of second-order diffusion tensors used HOTs to account for diffusion with generalized angular profiles while preserving its radial monoexponential behavior [86]. Even order Cartesian tensors were used to measure the apparent diffusivities (ADC) from the generalized Stejskal-Tanner equation as described in Eq. (2).

The simplest method [86] for estimating such tensors, $\mathscr{D}$, from the diffusion signal is to linearize the Stejskal-Tanner equation by taking the logarithm of Eq. (2). This leads to system of linear equations: $\mathbf{Ax} = \mathbf{y}$, where the rows of the design matrix $\mathbf{A}$ contain the monomials of the homogeneous form $D(\mathbf{u}) = \mathscr{D}^{(k)} \cdot^k \mathbf{u}$, the vector $\mathbf{y}$ contains the log-normalized diffusion signal scaled by the acquisition parameter b, and the vector $\mathbf{x}$ contains the unknown coefficients of the tensor $\mathscr{D}$. This system is overdetermined when the number of data acquisitions is greater than the number of unknown tensor coefficients and can be solved uniquely in the least squares sense by taking the Moore-Penrose pseudo-inverse of $\mathbf{A}$.

Since diffusivity is a non-negative physical quantity, the homogeneous form $D(\mathbf{u})$ cannot be negative for any $\mathbf{u} \in S^2$. This leads to a positivity constraint that needs to be respected while estimating $\mathscr{D}$. The least squares approach often violates this constraint for $\mathscr{D}$ with high orders and when the acquisitions are noisy.

Descoteaux et al. [28] proposed a linear approach with angular regularization to account for noisy acquisitions. Leveraging the bijection between HOTs and SHs, see Eq. (21), they estimated the coefficients of $\mathscr{D}$ by first estimating the coefficients in an SH basis of rank equal to the order of the tensor while applying Laplace-Beltrami smoothing on the sphere and then converting back to the tensor basis. This again leads to a linear system that is overdetermined when the number of acquisitions is larger than the number of tensor coefficients,

$$\mathbf{x} = \mathbf{M}^{-1}(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L})^{-1}\mathbf{B}^T\mathbf{y}, \tag{27}$$

where $\mathbf{x}$ contains the unique tensor coefficients, $\mathbf{y}$ contains the log-normalized signal, $\mathbf{B}$ is the design matrix in the SH basis and $\mathbf{M}$ represents the linear transformation matrix between the HOT basis and the SH basis. The matrix $\mathbf{L}$ is a diagonal matrix with entries $l_{ii} = \ell_i^2(\ell_i + 1)^2$, which represents the Laplace-Beltrami regularization of the SH Y_ℓ^m, and λ is the regularization weight. This becomes the least squares solution when $\lambda = 0$, but with nonzero λ, $\mathbf{L}$ tends to smooth higher order terms more, therefore, dampening the effects of noise in higher orders.

Florack et al. used the same Laplace-Beltrami regularization on the sphere, but for tensors instead of SHs [34]. This was based on an infinite inhomogeneous tensor basis representation, much like the SHs, with the diffusion function modified to $\tilde{D}(\mathbf{u}) = \sum_{k=0}^{\infty} \mathscr{D}^{(k)} \cdot^k \mathbf{u}$. It was shown that on the sphere, this representation was

redundant, and when truncated to a finite order, it represented the same diffusion function as in Eq. (2). The relation between the homogeneous and inhomogeneous tensor representation has been addressed rigorously in [8]. The estimation process was specifically crafted such that higher order tensors only captured the residual information not available in lower order tensors. This resulted in a "canonical" tensorial representation where the span of a tensor of fixed order k formed a degenerate eigenspace for the Laplace-Beltrami operator with eigenvalue $-k(k+1)$, exactly like the SHs.

The problem of estimating $\mathscr{D}$ with the positivity constraint was solved for order 4 tensors, in two different ways. The homogeneous forms of symmetric order 4 tensors of dimension 3 are known as ternary quartics. Barmpoutis et al. [9, 10] and Ghosh et al. [43] use Hilbert's theorem on positive semi-definite (psd) ternary quartics:

Theorem 1. *If $P(x, y, z)$ is homogeneous, of degree 4, with real coefficients and $P(x, y, z) \geq 0$ at every $(x, y, z) \in \mathbb{R}^3$, then there are quadratic homogeneous polynomials f, g, h with real coefficients, such that $P = f^2 + g^2 + h^2$.*

Therefore, estimating $P(x, y, z)$ (or $\mathscr{D}^{(4)}$) by estimating f, g, h ensures $\mathscr{D}^{(4)}$ to be psd. However, these quadratic polynomials can only be uniquely determined up to a 3D rotation and up to a sign. In other words, if the 6 coefficients of f, g, h each are written as column vectors $\mathbf{w}_f, \mathbf{w}_g, \mathbf{w}_h$, respectively, and a 6×3 matrix $\mathbf{W} = [\mathbf{w}_f, \mathbf{w}_g, \mathbf{w}_h]$ is constructed, then $P(x, y, z) = \mathbf{v}^T \mathbf{W}\mathbf{W}^T \mathbf{v}$, where $\mathbf{v}^T = [x^2, y^2, z^2, xy, xz, yz]$. Thus $\mathbf{W}$, $-\mathbf{W}$ and $\mathbf{WR}$ for any 3×3 orthogonal matrix $\mathbf{R}$ result in the same P.

Initially, Barmpoutis et al. fixed $\mathbf{R}$ by choosing the rotation that renders $\mathbf{A}$ – the top 3×3 block of $\mathbf{W}$ – to a lower triangular matrix [10]. This was achieved by considering the QR-decomposition of $\mathbf{A}$, but in practice $\mathbf{A}$ was taken to be lower triangular. This resulted in a reduction of unknown coefficients from $18 = 3 \times 6$ to 15, which is exactly the number of unique coefficients of $\mathscr{D}^{(4)}$. In a later work [9], an Iwasawa decomposition of $\mathbf{W}\mathbf{W}^T$ was taken, which implied the Cholesky decomposition of $\mathbf{A}$. This again resulted in $\mathbf{A}$ being rendered lower triangular – defining uniqueness over 3D rotations and again reducing the number of unknowns to 15. Furthermore, the Cholesky decomposition constrained the diagonal entries of $\mathbf{A}$ to be positive – defining uniqueness over the sign.

Ghosh et al. [43] estimated all 18 unknowns of $\mathbf{W}$ and reconstructed the 15 coefficients of $\mathscr{D}^{(4)}$ from the Gram matrix $\mathbf{W}\mathbf{W}^T$. Although $\mathbf{W}$ cannot be estimated uniquely, the Gram matrix representing the homogeneous form P is unique and the mapping from the coefficients of the Gram matrix to the coefficients of $\mathscr{D}^{(4)}$ is unique. Therefore, the estimation of the tensor coefficients is unambiguous. While Barmpoutis et al. [9, 10] used a Levenberg-Marquardt optimization scheme, Ghosh et al. [43] prefer the Broyden-Fletcher-Goldfarb-Shanno (BFGS) scheme.

Barmpoutis et al. [9, 10] further introduced an $L2$ distance measure between the homogeneous forms corresponding to the tensors evaluated on the unit sphere

$$\text{dist}(\mathscr{D}_1^{(4)}, \mathscr{D}_2^{(4)})^2 = \frac{1}{4\pi} \int_{S^2} [D_1(\mathbf{u}) - D_2(\mathbf{u})]^2 d\mathbf{u}, \tag{28}$$

which was computed analytically in terms of the difference of the coefficients of $\mathscr{D}_1^{(4)}$ and $\mathscr{D}_2^{(4)}$, and which was used for spatial regularization of the tensor field to account for noise.

A second way of estimating $\mathscr{D}^{(4)}$ with the positivity constraint was proposed by Ghosh et al. [44]. In this approach, the 6×6 isometrically equivalent matrix representation [20] $\mathbf{D}$ of $\mathscr{D}^{(4)}$ was used. Since $\mathbf{D}$ is symmetric and its positive definiteness ensures $\mathscr{D}^{(4)}$ to be positive, the affine invariant Riemannian metric for the space of symmetric positive definite matrices [71] was used to estimate $\mathbf{D}$ via a Riemannian gradient descent. However, the symmetry of the tensor $\mathscr{D}^{(4)}$ cannot be entirely captured by $\mathbf{D}$, which has 21 unique coefficients. Therefore, a final symmetrizing step was used to recover a positive and symmetric tensor $\mathscr{D}^{(4)}$.

The problem of estimating an arbitrary even order HOT, $\mathscr{D}^{(2k)}$, with the positivity constraint was also solved in two different ways. Barmpoutis et al. [13] used a result that states that for any even degree, $2k$, a (homogeneous) polynomial positive on the unit sphere can be written as a sum of squares of polynomials, p, of degree k on the unit sphere, $D(\mathbf{u}) = \mathscr{D}^{(2k)} \cdot^{(2k)} \mathbf{u} = \sum_{j=1}^{R} \lambda_j p^{(k)}(\mathbf{u}, \mathbf{c}_j)^2$, where λ_j are all positive and $\mathbf{c}_j$ are the coefficient vectors of the polynomials p_j with $||\mathbf{c}_j|| = 1$. However, since R, the number of polynomials in the sum, cannot be determined, they reformulated the problem as a spherical convolution problem $D(\mathbf{u}) = \int_{S^{\#\mathbf{c}-1}} \lambda(\mathbf{c}) p^{(k)}(\mathbf{u}; \mathbf{c})^2 d\mathbf{c}$, where the unit sphere $S^{\#\mathbf{c}-1}$ is embedded in $\mathbb{R}^{\#\mathbf{c}}$, with $\#\mathbf{c}$ being the number of elements in $\mathbf{c}$. The convolution was solved numerically by discretizing $S^{\#\mathbf{c}-1}$ finely and $\mathscr{D}^{(2k)}$ was estimated by solving the least squares problem for the unknowns λ_j

$$E = \sum_{i=1}^{N} \left(S_i / S_0 - \mathrm{e}^{-b \sum_{j=1}^{r} \lambda_j p(\mathbf{g}_i; \mathbf{c}_j)^2} \right)^2 \tag{29}$$

using a non-negative least squares (NNLS) to ensure that all $\lambda_j \geq 0$. Eq. (29) essentially overestimates R by r by discretizing the convolution, while the NNLS tends to compute a sparse solution ensuring that Eq. (29) does not overfit the signal.

A second method for estimating even order psd HOTs based on convex optimization was proposed by Qi et al. [95]. It was shown that the set of order $2k$ psd HOTs, $\mathscr{D}$, form a closed convex cone $\mathscr{C}$ in $\mathbb{R}^n$, where $\mathscr{D}$ has n unique coefficients and can be represented by $\mathbf{x} \in \mathbb{R}^n$. Furthermore, the psd constrained least squares estimation was shown to be convex and quadratic with a unique minimizer $\mathbf{x}^* \in \mathscr{C}$ such that if the unconstrained solution $\bar{\mathbf{x}} \in \mathbb{R}^n \setminus \mathscr{C}$ then $\mathbf{x}^* \in \partial\mathscr{C}$, the boundary of $\mathscr{C}$. The explicit psd constraint on $\mathscr{D}$ was formulated as $\lambda_{\min}(\mathscr{D}) \geq 0$ where $\lambda_{\min}(\mathscr{D})$, the minimum Z-eigenvalue of $\mathscr{D}$, was shown to be computationally tractable. The psd HOT $\mathscr{D}_{\mathbf{x}^*}$ (corresponding to $\mathbf{x}^*$) was estimated by first checking the psd-ness of the unconstrained HOT $\mathscr{D}_{\bar{\mathbf{x}}}$. If $\lambda_{\min}(\mathscr{D}_{\bar{\mathbf{x}}}) \geq 0$, then by uniqueness $\mathscr{D}_{\mathbf{x}^*} = \mathscr{D}_{\bar{\mathbf{x}}}$. However, if $\bar{\mathbf{x}} \notin \mathscr{C}$, then $\mathscr{D}_{\mathbf{x}^*}$ was estimated by solving the non-differentiable, *non-convex* optimization problem $L(\mathbf{x}) = \min\{|\mathbf{A}\mathbf{x} - \mathbf{y}|^2 : \lambda_{\min}(\mathscr{D}_{\mathbf{x}}) = 0\}$, with only an equality constraint, by a subgradient descent approach. In theory, $\mathscr{D}_{\mathbf{x}^*}$ could also be estimated by solving the psd constrained non-differentiable *convex* least squares problem.

Alternatively, Barmpoutis et al. [11] used even ordered HOTs to model the logarithm of the diffusivities. This preserved the monoexponential radial diffusion but considered the exponential of the tensor for the angular diffusion $D(\mathbf{u}) = \exp(\mathscr{D}^{(k)} \cdot^k \mathbf{u})$ (in Eq. (2)). This automatically ensured positive diffusion without having to impose any constraints. The approach was inspired by the Log-Euclidean metric for DTI [3].

4.2 Fitting Models of Apparent Diffusional Kurtosis

Fitting the coefficients of the diffusion tensor $\mathbf{D}$ and kurtosis tensor $\mathscr{W}$ in Eq. (5) is simplified by initially considering each gradient direction separately, and finding parameters of the corresponding one-dimensional diffusion process,

$$S(b) = S_0 \mathrm{e}^{-bd + \frac{1}{6}(bd)^2 K}, \tag{30}$$

where d and K are apparent diffusion and kurtosis coefficients, respectively. Estimating these two variables requires measurements $S(b)$ with at least two non-zero b-values, in addition to the baseline S_0 measurement. After taking the logarithm on both sides of Eq. (30), this leads to a system of equations that is quadratic in d, and can thus no longer be solved with a linear least squares estimator. Instead, gradient-based iterative Levenberg-Marquardt optimization has been employed [61].

Assuming a Gaussian noise model results in a positive bias in the estimated kurtosis values, which can be removed by finding the maximum likelihood fit under a Rician noise model [111] or, more easily, by accounting for the noise-induced bias in the measurements themselves [61, 82]. This is done by adding an estimate η of the background noise to the signal model in Eq. (30),

$$S(b) = \sqrt{\eta^2 + \left(S_0 \mathrm{e}^{-bd + \frac{1}{6}(bd)^2 K}\right)^2}. \tag{31}$$

After finding parameters d_i and K_i for each individual gradient direction i, a second-order diffusion tensor $\mathbf{D}$ can be fit linearly to the d_i. Given this estimate of $\mathbf{D}$, the fourth-order kurtosis tensor $\mathscr{W}$ can then be fit linearly using Eq. (4) [69, 82].

Kurtosis is a dimensionless quantity and can, in theory, take on any value $K \geq -2$. However, the kurtosis of a system that contains noninteracting Gaussian compartments with different diffusivities is always non-negative, and empirical results suggest non-negative kurtosis in human brain tissue [61]. Similarly, an upper bound on kurtosis, $K_i \leq 3/(b_{\max} d_i)$, where $b_{\max}$ is the largest b-value used in the measurements, is implied by the empirical observation that in practice, the signal $S(b)$ is a monotonically decreasing function of b. These two constraints have been enforced as part of the fitting, using quadratic programming or heuristic thresholding

[105]. Other authors have chosen to merely enforce the lower bound $K \geq -3/7$, which correspond to the kurtosis of water confined to equally-sized spherical pores, by a sum-of-squares parametrization of the homogeneous polynomial represented by $\mathscr{W}$ [16]. Additional regularization has been employed to penalize extrema in the homogeneous form that fall outside the range of the measured kurtosis values [65].

4.3 Fitting Deconvolution-Based Models

Spherical deconvolution models the diffusion-weighted signal $S(\mathbf{u})$ in different gradient directions $\mathbf{u}$ as the convolution of a fiber orientation density function (fODF) F with a response function R. It describes the signal attenuation caused by a single nerve fiber bundle, and it is assumed to be cylindrically symmetric:

$$S(\mathbf{u}) = \iint_{\|\mathbf{v}\|=1} F(\mathbf{v})\ R(\mathbf{v}\cdot\mathbf{u})\ d\,\mathbf{v} \tag{32}$$

Based on Eq. (32), deconvolution can be used to estimate the fiber ODF F from the measurements S. Deconvolution is done most easily in the spherical harmonics basis, where it amounts to simple scalar division. However, constructing a spherical harmonics representation of the deconvolution kernel R requires two choices: Beside estimating the response of a single fiber compartment from the data [107] or deriving it from an analytical fiber model [30, 101], it involves deciding how the single fiber compartment should be represented *after* deconvolution [106].

Even though the delta distribution may seem like an obvious choice, it requires an infinite number of coefficients in the spherical harmonics basis. Therefore, Tournier et al. [106] approximate the delta peak, resulting in non-trivial interactions between peaks of non-orthogonal fiber compartments and leading to systematic errors when taking ODF maxima as estimates of fiber directions, even when no measurement noise is present. Schultz and Seidel [100] have removed this problem by instead modeling single fiber peaks as rank-1 tensors, and performing a low-rank approximation of the resulting order-p fODF tensor $\mathscr{F}$,

$$\underset{\lambda_i,\mathbf{v}_i}{\operatorname{argmin}} \left\| \mathscr{F} - \sum\nolimits_{i=1}^{r} \lambda_i \mathbf{v}_i^{\otimes p} \right\|_F, \tag{33}$$

where $\mathbf{v}_i$ describe the per-compartment principal directions, and λ_i are proportional to their volume fractions. The approximation rank r corresponds to the number of discrete fiber compartments; one way to estimate it is by learning from simulated training data via support vector regression [97].

This tensor-based variant of spherical deconvolution uses the linear bijection between spherical harmonics and polynomial bases (cf. Sect. 3.3) twice: First, to map a rank-1 tensor of the same order as the desired fODF tensor $\mathscr{F}$ to the spherical harmonics basis, which is required to find the correct kernel $\mathbf{R}$ for use with that

tensor order. Second, to transform the deconvolution result F, obtained in the spherical harmonics basis, back into its tensor representation $\mathscr{F}$.

Since compartments cannot have negative weights, valid fODF tensors should permit a positive decomposition into rank-1 terms. For tensor order $k > 2$, this is a stronger requirement than non-negativity of the homogeneous form, which is a more natural constraint for models of apparent diffusivity (Sect. 4.1). It can be enforced by computing an approximation with the generic number of rank-1 terms and non-negative weights [98].

Similar to a previous approach of Barmpoutis et al. [13], Weldeselassie et al. [113] enforce non-negativity of F by parametrizing the homogeneous polynomial $\mathscr{F}$ (with even order k) as a sum of squares of polynomials of order $k/2$. Rather than performing the deconvolution in spherical harmonics, they discretize the fODF, so that it can be found as the non-negative least squares solution of a linear system.

4.4 Fitting Other Types of Models

When fitting the higher-order diffusion model described by Eq. (3) [77,78], we only consider tensors of even order, as was argued in Sect. 2. By taking the logarithm and truncating after order $2n$, the equation can be rewritten in the form

$$\mathrm{Re}[\log(S(\mathscr{B})/S_0)] = \sum\nolimits_{k=1}^{n} (-1)^k \left\langle \mathscr{B}^{(2k)}, \mathscr{D}^{(2k)} \right\rangle,$$

where Re denotes the real part of the logarithmic signal and the inner products between tensors $\mathscr{B}^{(2k)}$ and $\mathscr{D}^{(2k)}$ is defined in Eq. (16). Tensors $\mathscr{D}^{(2k)}$ can be estimated by considering measurements with different gradient strengths and directions, which lead to different $\mathscr{B}_i^{(2k)}$, and truncating the tensor series at the desired order. If we have m measurements, we obtain m equations of the above form, linear in the coefficients of $\mathscr{D}$. These can be combined in a matrix equation

$$\mathbf{y}(\log(|S_i(\mathscr{B}_i)|/|S_0|)) = \mathbf{B}(\mathscr{B}^{(2k),i})\,\mathbf{x}(\mathscr{D}^{(2k)}), \tag{34}$$

where $i = 1, \ldots, m$. In practice, the modulus $|\cdot|$ rather than the real part of the complex signal is used, since phase is unreliable. The vector $\mathbf{x}$, which contains the coefficients of $\mathscr{D}^{(2k)}$, can be estimated by solving Eq. (34) in the least squares sense.

Higher-order tensors representing q-ball ODFs (see Sect. 2) can also be fitted to HARDI data. An analytical solution for the q-ball ODF is given by Anderson [2], Hess et al. [50], and Descoteaux et al. [29]

$$\psi_{\text{q-ball}}(\mathbf{u}) = \sum\nolimits_{i=1}^{N} 2\pi P_{l_i}(0)\, c_i\, Y_i(\mathbf{u}) \tag{35}$$

where $\mathbf{u}$ is a unit norm vector, P_{l_i} is the Legendre polynomial of degree l_i, $\{Y_i\}_{i=1}^{N}$ is a modified SH basis as in Eq. (18), and c_i are the harmonic coefficients of the MR

signal. A tensor representation of $\psi_{\text{q-ball}}$ can be obtained from the bijection between SHs and tensors. Alternatively, it can be reconstructed directly in a tensor basis [34]

$$\psi_{\text{q-ball}}(\mathbf{u}) = \sum_{k=0}^{n} 2\pi P_{l_k}(0) \mathcal{S}_k \cdot^k \mathbf{u} \tag{36}$$

where n is the maximum order of a series of tensors $\mathcal{S}_k$ fitted to the diffusion signal such that higher orders only encode the fitting residuals from lower orders.

5 Processing Higher-Order Tensors in Diffusion MRI

5.1 Computing Rotationally Invariant Scalar Measures

It is desirable to extract meaningful scalars from the estimated higher-order tensors. In particular, rotationally invariant quantities are preferable. These are independent of the coordinate system and thus intrinsic features of the tensor.

5.1.1 Higher-Order Diffusion Tensors

Rotationally invariant measures of diffusivity and anisotropy based on higher-order diffusion tensors have been proposed in [89]. The mean diffusivity is defined as:

$$\langle D \rangle = \frac{1}{4\pi} \iint_{S^2} D(\mathbf{u})\, d\mathbf{u} \tag{37}$$

where $\mathbf{u}$ is a unit direction vector and $D(\mathbf{u})$ are the diffusivities as in Eq. (2). The generalized anisotropy (GA) and scaled entropy (SE) are given by

$$\text{GA} = 1 - \frac{1}{1 + (250\text{V})^{\varepsilon(V)}} \quad \text{and} \quad \text{SE} = 1 - \frac{1}{1 + (60(\ln 3 - \eta))^{\varepsilon(\ln 3 - \eta)}}, \tag{38}$$

where $\varepsilon(\gamma) = 1 + 1/(1 + 5{,}000 \cdot \gamma)$ and V and η are the variance and entropy of the normalized diffusivities, $D(\mathbf{u})/3\langle D \rangle$. The definition of these measures does not rely on any specific tensor order. In addition, GA and SE are scaled between 0 and 1. Note that these measures can also be calculated from other functions defined on the unit sphere, such as orientation distribution functions.

GA and SE values for simulated data modelling two and three fibers show a clear difference between those implied by tensors of order 2 and higher-order (4, 6 and 8) tensors, the latter being significantly higher [27, 28, 84, 89]. GA and SE have also been reported to be slightly higher in the case of sixth order tensors than for order 4 [27]. On the other hand, for data simulating one fiber, GA and SE are independent of the tensor order. This is also the case for the mean diffusivity [89].

Table 1 Mean kurtosis and kurtosis anisotropy. β: D-eigenvalue of $\mathscr{W}$, ν: number of D-eigenvalues, N: total number of diffusion measurement directions, K_{app}: AKC in a particular direction as in Eq. (5), $\mathbf{e_i}$ ($i = 1, 2, 3$): eigenvector of diffusion tensor $\mathscr{D}$, $K_{\text{app}}(\mathbf{e_i}) = (\text{MD}^2/\lambda_i^2) \cdot \hat{\mathscr{W}}_{iiii}$, $\hat{\mathscr{W}}$: kurtosis tensor in the basis $\{\mathbf{e_i}\}$, $\bar{K} = (1/3)(K_{\text{app}}(\mathbf{e_1}) + K_{\text{app}}(\mathbf{e_2}) + K_{\text{app}}(\mathbf{e_3}))$

Reference	MK	FA_K
[94]	$\frac{(\text{MD})^2}{\nu} \sum_{i=1}^{\nu} \beta_i$	$\sqrt{\frac{\nu}{\nu-1}} \sqrt{\frac{\sum_{i=1}^{\nu} (\beta_i - \text{MK}/(\text{MD})^2)^2}{\sum_{i=1}^{\nu} \beta_i^2}}$
[56]	$\frac{1}{N} \sum_{i=1}^{N} K_{\text{app}}(\mathbf{u}_i)$	$\sqrt{\frac{3}{2}} \sqrt{\frac{\sum_{i=1}^{3} (K_{\text{app}}(\mathbf{e_i}) - \bar{K})^2}{\sum_{i=1}^{3} K_{\text{app}}^2(\mathbf{e_i})}}$
[60, 93, 105] (MK) [91]	$\frac{1}{4\pi} \iint_{S^2} K_{\text{app}}(\mathbf{u})\, d\mathbf{u}$	$\sqrt{\frac{1}{4\pi} \iint_{S^2} (K_{\text{app}}(\mathbf{u}) - \text{MK})^2\, d\mathbf{u}}$
[81]	Idem	$\sqrt{\frac{\iint_{S^2} (K_{\text{app}}(\mathbf{u}) - \text{MK})^2\, d\mathbf{u}}{\iint_{S^2} K_{\text{app}}^2(\mathbf{u})\, d\mathbf{u}}}$

Table 2 Axial and radial kurtoses. $\mathbf{e_\phi} = (0, \cos\phi, \sin\phi)$ in the basis $\{\mathbf{e_i}\}$

Reference	$K_{\parallel}$	$K_{\perp}$
[56]	$K_{\text{app}}(\mathbf{e_1})$	$\frac{K_{\text{app}}(\mathbf{e_2}) + K_{\text{app}}(\mathbf{e_3})}{2}$
[91] ($K_{\perp}$) [60, 105]	Idem	$\int_0^{2\pi} K_{\text{app}}(\mathbf{e_\phi})\, d\phi$

GA and SE for real HARDI data of healthy subjects have been studied in [27, 83, 84]. It has been shown that fourth- and sixth-order tensors result in increased values for both measures, especially for SE, with respect to second-order tensors. This effect is observed in areas with intra-voxel orientational heterogeneity but also in some regions with coherent axonal orientation. On the other hand, GA and SE become more sensitive to noise for increasing tensor order [27].

The variance of fourth-order covariance tensors has also been investigated for DTI data of glioblastoma patients [32]. Results indicate a better variance contrast between tumor subregions than for FA.

5.1.2 Diffusional Kurtosis Tensors

A number of rotationally invariant scalar measures based on fourth-order kurtosis tensors have been proposed. Different definitions of mean kurtosis (also referred to as average AKC), kurtosis anisotropy, radial and axial kurtoses can be found in the literature. Some of them are related to certain eigenvalues of the kurtosis tensor, which we discuss later in this section. These measures are summarized in Tables 1 and 2. It is clear that they are rotationally invariant, since both the AKC and eigenvalues involved in their definition are rotationally invariant.

Note that the first two definitions of kurtosis anisotropy in Table 1 are completely analogous to the DTI case but based on the kurtosis tensor D-eigenvalues and AKC values along the diffusion tensor eigenvectors, respectively. As FA, FA_K takes on values $0 \leq \text{FA}_K \leq 1$, except for the definition in [91].

Some of these measures have been probed for in vivo and ex vivo rat brain DKI, and compared to their DTI analogues [56]. Mean and radial kurtoses showed strong contrast between GM and WM both in and ex vivo. In particular, radial

kurtosis performs better than all other directional diffusivities and kurtoses. For axial kurtosis, a stronger contrast was observed under ex vivo conditions. On the other hand, kurtosis anisotropy was similar to FA both in and ex vivo.

Mean kurtosis and kurtosis anisotropy have also been computed by an adaptive spherical integral, and compared to those based on D-eigenvalues for real diffusion data of a healthy subject and a stroke patient [81]. The latter are seen to be more sensitive to noise. Exact expressions for mean and radial kurtoses can be obtained [105]. These have been shown, together with axial kurtosis, on DKI scans of healthy subjects [60, 105]. The optimization of the diffusion gradient settings for estimation of mean and radial kurtosis, and kurtosis anisotropy has been studied as well. It has been shown that this increases precision considerably [91].

D-eigenvalues of the fourth-order kurtosis tensor $\mathscr{W}$ are defined by Qi et al. [94]

$$\mathscr{W} \cdot^3 \mathbf{x} = \beta\, \mathbf{D}\mathbf{x}; \quad \mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{x} = 1, \tag{39}$$

where $\mathbf{x}$ is the D-eigenvector associated to D-eigenvalue β. D-eigenvalues have been shown to be rotationally invariant [94]. The largest and smallest D-eigenvalues can be used to compute the largest and smallest AKC values as $(\mathrm{MD})^2\beta_{\max}$ and $(\mathrm{MD})^2\beta_{\min}$. Other type of eigenvalues which have been studied in this context are the Kelvin eigenvalues of the kurtosis tensor, which are also rotationally invariant. A three-dimensional symmetric fourth-order tensor can be mapped to a six-dimensional second-order tensor. The eigenvalues $(\eta_1, \ldots, \eta_6)$ of its matrix representation, a symmetric 6×6 matrix, are the Kelvin eigenvalues of the considered fourth-order tensor. It has been shown that the largest and smallest Kelvin eigenvalues of (a scaled version of) the kurtosis tensor $\hat{\mathscr{W}}$ are, respectively, an upper and lower bound of the largest and smallest AKC values [93]. The interpretation of Kelvin eigenvalues in terms of AKC values is thus less clear than for D-eigenvalues.

5.1.3 Orientation Distribution Functions

ODF maxima are characterized by their position and value (see Sect. 5.3), but also by their geometric shape. A peak sharpness measure

$$\mathrm{PS} = \frac{-\mu_1}{k\, F(\mathbf{u})} \tag{40}$$

can be derived from the value $F(\mathbf{u})$, order k of $\mathscr{F}$, and a Hessian eigenvalue μ_1 of F (at maxima, $\mu_2 \leq \mu_1 \leq 0$). The homogeneous forms of second-order tensors $\mathbf{F}$ have a single maximum, whose sharpness depends on the degree to which $\mathbf{F}$ has a linear shape, as measured by the widely used invariant $c_l = (\lambda_1 - \lambda_2)/\lambda_1$ [114]. In fact, when applied to a second-order tensor, $\mathrm{PS} = c_l$ [98].

Peak Fractional Anisotropy (PFA) is designed to coincide with traditional Fractional Anisotropy (FA) [21] when the diffusion process is well-described by

a second-order diffusion tensor, but generalizes it to a per-peak measure in case of more than one ODF maximum [41]. It is defined by fitting a second-order tensor to each ODF peak and computing its FA. Based on the function value F and principal curvatures $\kappa_1 > \kappa_2$ at the maximum, the fitted tensor eigenvalues are given by:

$$\text{ODF-T}: \quad \lambda_1 = F^2,\ \lambda_2 = \frac{F}{\kappa_2},\ \lambda_3 = \frac{F}{\kappa_1} \tag{41}$$

$$\text{ODF-SA}: \lambda_1 = 1,\ \lambda_2 = \frac{3}{2+\kappa_2 F},\ \lambda_3 = \frac{3}{2+\kappa_1 F} \tag{42}$$

ODF-T refers to the q-ball defined by Tuch [109]; ODF-SA denotes a solid angle ODF [1, 108]. The total PFA is defined by considering a weighted sum of the PFA over all ODF maxima:

$$\text{Total-PFA} = \sum_{i=1}^{\sharp_{\max.}} F_i \cdot \text{PFA}_i \tag{43}$$

Unlike Fractional Anisotropy, Total-PFA is able to distinguish between near-isotropic regions with many weak ODF maxima and areas with complex fiber structure, which exhibit multiple, high anisotropy maxima.

Other geometrical scalars have also been considered. The Ricci scalar is a well-known invariant quantity in differential geometry representing intrinsic curvature, and constructed from the metric and metric-derived tensors. It has been proposed as a DTI scalar measure in the context of Riemannian geometry [35]. The Ricci scalar can also be calculated from a (strongly) convexified ODF by relating it to Finsler geometry (see Sect. 5.5 and chapter "Riemann-Finsler Geometry for Diffusion Weighted Magnetic Resonance Imaging") [6]. However, experimental results on the latter have not yet been reported.

In addition, principal invariants of fully symmetric fourth-order tensors representing an ODF have been studied [36]. Invariants of fourth-order covariance tensors in DTI had been previously investigated [20]. More general invariants of fourth-order tensors have been recently presented [42]. Principal invariants can be computed from the tensor Kelvin eigenvalues $(\eta_1, \ldots, \eta_6)$ (see Sect. 5.1.2):

$$\begin{aligned}
I_1 &= \eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5 + \eta_6 \\
I_2 &= \eta_1\eta_2 + \eta_1\eta_3 + \ldots + \eta_5\eta_6 \\
I_3 &= \eta_1\eta_2\eta_3 + \eta_1\eta_2\eta_4 + \ldots + \eta_4\eta_5\eta_6 \\
I_4 &= \eta_1\eta_2\eta_3\eta_4 + \eta_1\eta_2\eta_3\eta_5 + \ldots + \eta_3\eta_4\eta_5\eta_6 \\
I_5 &= \eta_1\eta_2\eta_3\eta_4\eta_5 + \ldots + \eta_2\eta_3\eta_4\eta_5\eta_6 \\
I_6 &= \eta_1\eta_2\eta_3\eta_4\eta_5\eta_6
\end{aligned} \tag{44}$$

These quantities are, by definition, rotationally invariant and can therefore be used as building blocks for invariant scalar HARDI measures. Experiments on HARDI phantom and brain data have been presented but further work is required to asses the

utility of principal invariants in this context. Finally, note that both the Ricci scalar and principal invariants can also be calculated from higher-order diffusion tensors.

5.2 *Reconstructing the Diffusion Propagator*

The diffusion process is characterized by a probability density function $P(\mathbf{r}, t)$ that specifies the probability of a spin displacement $\mathbf{r}$ within diffusion time t. $P(\mathbf{r}, t)$ is known as the diffusion propagator or Ensemble Average Propagator (EAP). It is related to the dMRI signal by a Fourier transform in the q-space formalism $S(\mathbf{q}, t)/S_0 = \int_{\mathbb{R}^3} P(\mathbf{r}, t) \mathrm{e}^{2\pi i \mathbf{q}\cdot\mathbf{r}} d\mathbf{r}$ [25]. Even though higher order tensor estimates of ADC and kurtosis can discern regions with multiple fiber directions, they cannot be used to resolve the directions themselves. To resolve fiber directions the EAP or its characteristics such as the ODF need to be computed.

In DTI, the diffusivities are modeled by a quadratic function given by the diffusion tensor, Eq. (1). The Fourier transform of the resulting signal yields the corresponding EAP, an oriented Gaussian distribution with the tensor's largest eigen-pair indicating the single major fiber direction. However, when HOTs are used to model more complex ADC profiles, computing the EAP turns out to be a trickier problem.

Unlike in DTI, the analytical Fourier transform of the tensor model in Eq. (2) is unknown. In [88], a fast Fourier transform was performed on interpolated (and extrapolated) q-space data on a Cartesian grid generated from the tensor in Eq. (2) to numerically estimate the EAP. In [87], an analytical EAP on a single R_0-shell, i.e., $P(R_0 \frac{\mathbf{r}}{||\mathbf{r}||})$, was proposed for this model. However, in this Diffusion Orientation Transform (DOT), the SH basis representation of the tensor was used, see Eq. (21).

In [40], the authors considered a modified non-monoexponential model inspired from Eq. (2) where the HOT was used to describe the signal in the entire q-space. The modified model leads to an analytical series expansion of the EAP in Hermite polynomials. In [15], the authors proposed to use tensors to describe a single R_0-shell of the EAP, $P(R_0 \frac{\mathbf{r}}{||\mathbf{r}||})$. They used Hermite polynomials to describe the dMRI signal, since under certain constraints the Fourier transform of Hermite polynomials are homogeneous forms or tensors. Note that [40] and [15] used the same dual Fourier bases but in the opposite spaces to analytically resolve the Fourier transform.

The first attempt to estimate the EAP analytically was based on the tensor model in Eq. (3), where the HOTs represented the cumulant tensors of the EAP since the dMRI signal is also the characteristic function of the EAP. The authors in [77, 78], proposed to use the Gram-Charlier series to compute a series estimate of the EAP from the first four cumulant tensors, i.e., covariance (diffusion) and kurtosis. In theory, the Gram-Charlier series could be improved by the Edgeworth series [45].

In [69], the authors computed the ODF directly from the first four cumulant tensors – diffusion and kurtosis. In contrast to [77, 78], they do not estimate the full EAP, but only its radial marginalization.

5.3 Finding Maxima of the Homogeneous Form

The maxima of many orientation distribution functions in dMRI, which can be represented in the HOT or SH bases, indicate underlying fiber directions. It is, therefore, crucial to compute these maxima with high precision.

The simplest approach is to discretely sample the homogeneous form on a spherical mesh and to compare its values on the finite vertices to approximately identify the maxima [54]. However, even a 16th order tessellation of the icosahedron or 1,281 vertices on the sphere can lead to an error of $\sim 4°$. Numerical optimization techniques such as Newton-Raphson and Powell's methods have been used in the SH basis [58, 107] to overcome this limitation. In [55], numerical optimization was combined with the Euler integration step of a tractography algorithm in the tensor basis to trace fibers efficiently.

However, such local optimization techniques are highly dependent on initialization. In [23] and [47] two methods were shown for computing all the stationary points of a homogeneous form. In [23], the Z-eigenvalue/eigenvector formulation was used and a system of two polynomials in two variables – the homogeneous form and the unit sphere constraint – was solved using resultants (detailed in [95]). The stationary points were then classified by their principal curvatures into maxima, minima and saddle-points. In [46], the gradient of the homogeneous form constrained to the unit sphere – a system of four polynomials – was equated to zero. The roots of the system were computed by the subdivision method which ensures that all roots are analytically bracketed thus missing none. The stationary points were then classified into maxima, minima and saddle points using the Bordered Hessian.

5.4 Applications of Tensor Decompositions and Approximations

There are four lines of work that have applied tensor decompositions in the context of diffusion MRI. The first results from considering normal distributions of second-order diffusion tensors, which involve a fourth-order covariance tensor Σ. When the diffusion tensor is written as a vector, Σ is naturally represented by a 6×6 symmetric positive definite matrix $\mathbf{S}$, to which the spectral decomposition into eigenvalues and eigentensors can be applied, in order to facilitate visualization and quantitative analysis [20]. Alternatively, Σ can be expressed in a local coordinate frame that is derived from invariant gradients and rotation tangents [63]. The coordinates in this frame isolate physically and biologically meaningful components such as variability that can be attributed to changes in trace, anisotropy, or orientation.

Second, the distribution of fiber orientation estimates, either from the diffusion tensor or from HARDI, has been modeled by mapping the corresponding probability measure into a reproducing kernel Hilbert space. With a power-of-cosine kernel, this results in a higher-order tensor representation, which can be decomposed into

a rank-1 approximation and a non-negative residual to visually and quantitatively investigate the uncertainty in fiber estimates from diffusion MRI [99].

Third, in the framework described in detail in Sect. 4.3, a low-rank approximation of fODF tensors provides a less biased estimate of principal directions than fODF maxima. It has been shown [101] that this model can be used to approximate and to more efficiently and robustly fit the ball-and-multi-stick model [22]. Subsequent work has imposed an additional non-negativity constraint during deconvolution, and proposed an alternative optimization algorithm [62]. Low-rank approximations were shown to produce useful estimates of crossing fibers even from a relatively small number of gradient directions [49].

Finally, another line of work has attempted to decompose higher-order diffusion tensors in order to obtain crossing fiber directions [59, 115]. However, these techniques are yet to be validated on synthetic data with varying crossing angles, and have not yet been shown to reconstruct known fiber crossings in real data.

5.5 Finslerian Tractography

DTI streamline tracking can be generalized to HARDI by means of Finsler geometry. A second-order Finsler metric tensor can be defined at each point $\mathbf{q}$ from an ODF in the following way [4,5,7,34]

$$\hat{F}(\mathbf{q},\mathbf{x}) = \left(\sum\nolimits_{i_i \ldots i_p} \mathscr{F}_{i_1 \ldots i_p}(\mathbf{q}) x^{i_1} \ldots x^{i_p}\right)^{1/p}, \quad g_{ij}(\mathbf{q},\mathbf{x}) = \frac{1}{2}\frac{\partial \hat{F}^2(\mathbf{q},\mathbf{x})}{\partial x^i \partial x^j} \tag{45}$$

where $\mathscr{F}$ is an ODF tensor of (even) order p, $\hat{F}$ is the Finsler function and g_{ij} is the Finsler metric, $i, j = 1, \ldots, 3$, which depends on both position and direction. Note that this definition of the Finsler function $\hat{F}$ is by no means unique. In fact, this is still a subject of intensive research (see chapter "Riemann-Finsler Geometry for Diffusion Weighted Magnetic Resonance Imaging"). Thus a local diffusion tensor can be obtained per direction. Tracking can be performed by extracting the principal eigenvector of the diffusion tensor corresponding to the arrival direction. As long as this direction is sufficiently aligned to the eigenvector, and the diffusion tensor FA is above a certain treshold, tracking continues. Experiments on Finsler streamline tracking using fourth-order tensors have been presented on simulated fiber crossings and real HARDI data. It has been shown that Finsler streamlines can, unlike DTI streamlines, correctly cope with nerve fiber bundle crossings.

5.6 Registration and Atlas Construction

Registration transforms data sets from different times or subjects to a common coordinate system, so that anatomical structures align. Atlas construction is based

on registering a large number of subjects, in order to obtain a description of average anatomy, and of the most common modes of variation. Modeling parameters of the diffusion process with higher-order tensors makes registration of tensor fields a relevant research problem. Registration requires selection of an appropriate metric to measure the dissimilarity between individual tensors; for this purpose, Barmpoutis et al. [12, 14] propose two alternative choices, which are both scale and rotation invariant. Integrating the local dissimilarity over the domain of the tensor field results in an overall measure of dissimilarity. Registration is achieved by finding the coordinate transformation that minimizes this measure.

It is important to also transform the individual tensors according to the coordinate transformation applied to the domain of the field. For example, when the domain of the tensor field is rotated, a corresponding rotation of the tensors themselves is required in order to preserve relevant structures, such as the trajectories of nerve fiber bundles. When the transformation is (locally) affine, it has been proposed to simply apply it to the tensors via Eq. (11) [14]. Alternative methods for transformation have been proposed based on the spectral decomposition [96] and different sum-of-squares parametrizations [9, 48, 96].

6 Conclusion

The wide range of models and computational methods that have been surveyed in this chapter testify to the power and flexibility that higher-order tensors provide for the analysis of data from diffusion MRI, and to the increasing momentum of the research associated with this topic. Generalized eigenvalues, scalar invariants, tensor decompositions, and low-rank approximations have all proven valuable in the context of this application.

Looking ahead, several *theoretical problems* remain to be solved. While many approaches have focused on the properties of individual tensors, less attention has been paid to the *global nature of the tensor fields* that arise in diffusion MRI. The recent use of Finsler geometry is a natural step in this direction.

Even though *low-rank approximations* have proven to work well in practice, uniqueness of approximations over the reals is mostly open (for the complex case, see [66]). Moreover, we are still lacking algorithms with provable convergence properties, and formal results on the well-conditionedness of such approximations.

Many approaches have been proposed to ensure *non-negativity of higher-order tensors* that model apparent diffusivities (cf. Sect. 4.1). Less attention has been paid to the fitting of deconvolution models, which are constrained to the convex cone of tensors that can be expressed as a positive sum of rank-1 tensors; in general, that is a stricter constraint than non-negativity.

While many neuroscientific studies that use diffusion imaging are now published each month, they still almost exclusively use either the second-order diffusion tensor [21] or the ball-and-stick model [22]. A challenge in the next few years will be to

take approaches based on higher-order tensors into the application domain. This will require more work on several subproblems:

Statistical tests on *scalar invariants* such as Mean Diffusivity or Fractional Anisotropy are a mainstay of DTI-based studies. Even though a considerable number of invariants have now been derived from higher-order tensors (cf. Sect. 5.1), the practical utility of many of them is limited by their unclear biological or neuroanatomical interpretation.

Given an ever-increasing palette of models, it becomes a more urgent problem to pick one of them to test a given hypothesis, and to choose values for parameters such as tensor order, approximation rank, or regularization weights. Improved understanding of *formal relationships between different models* and mathematical rules for *model selection* are required.

Spatial coherence and *signal sparsity* need to be exploited in order to reliably estimate the large number of parameters in advanced models such as the ensemble average propagator, without requiring excessively time consuming measurements.

Acknowledgements A. Ghosh and R. Deriche are partially supported by the NucleiPark research project (ANR Program "Maladies Neurologique et maladies Psychiatriques") and the France Parkinson Association. L.-H. Lim is partially supported by an AFOSR Young Investigator Award (FA9550-13-1-0133), an NSF CAREER Award (DMS-1057064), and an NSF Collaborative Research Grant (DMS-1209136).

References

1. Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N.: Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle. Magn. Reson. Med. **64**(2), 554–566 (2010)
2. Anderson, A.W.: Measurement of fiber orientation distributions using high angular resolution diffusion imaging. Magn. Reson. Med. **54**(5), 1194–1206 (2005)
3. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magn. Reson. Med. **56**(2), 411–421 (2006)
4. Astola, L., Florack, L.: Finsler geometry on higher order tensor fields and applications to high angular resolution diffusion imaging. In: Proceedings of the International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), Voss, pp. 224–234. Springer (2009)
5. Astola, L., Florack, L.: Finsler geometry on higher order tensor fields and applications to high angular resolution diffusion imaging. Int. J. Comput Vis. **92**, 325–336 (2011)
6. Astola, L., Fuster, A., Florack, L.: A Riemannian scalar measure for diffusion tensor images. Pattern Recognit **44**, 1885–1891 (2011)
7. Astola, L., Jalba, A., Balmashnova, E., Florack, L.: Finsler streamline tracking with single tensor orientation distribution function for high angular resolution diffusion imaging. J. Math. Imaging Vis. **41**, 170–181 (2011)
8. Balmashnova, E., Fuster, A., Florack, L.: Decomposition of higher-order homogeneous tensors and applications to HARDI. In: Panagiotaki, E., O'Donnell, L., Schultz, T., Zhang, G.H. (eds.) Proceedings of the Computational Diffusion MRI (CDMRI), Nice, pp. 79–89 (2012)

9. Barmpoutis, A., Hwang, M.S., Howland, D., Forder, J.R., Vemuri, B.C.: Regularized positive-definite fourth order tensor field estimation from DW-MRI. NeuroImage **45**(1, Suppl. 1), S153–S162 (2009)
10. Barmpoutis, A., Jian, B., Vemuri, B.C., Shepherd, T.M.: Symmetric positive 4th order tensors & their estimation from diffusion weighted MRI. In: Karssemeijer, N., Lelieveldt B. (eds.) IPMI, Kerkrade. LNCS, vol. 4584, pp. 308–319 (2007)
11. Barmpoutis, A., Vemuri, B.C.: Exponential tensors: a framework for efficient higher-order DT-MRI computations. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Washington, DC, pp. 792–795 (2007)
12. Barmpoutis, A., Vemuri, B.C.: Groupwise registration and atlas construction of 4th-order tensor fields using the $\mathbb{R}^+$ Riemannian metric. In: Yang, G.Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI), Part I, London. LNCS, vol. 5761, pp. 640–647 (2009)
13. Barmpoutis, A., Vemuri, B.C.: A unified framework for estimating diffusion tensors of any order with symmetric positive-definite constraints. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Rotterdam, pp. 1385–1388 (2010)
14. Barmpoutis, A., Vemuri, B.C., Forder, J.R.: Registration of high angular resolution diffusion MRI images using 4th order tensors. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Part I, Brisbane. LNCS, vol. 4791, pp. 908–915 (2007)
15. Barmpoutis, A., Vemuri, B.C., Forder, J.R.: Fast displacement probability profile approximation from HARDI using 4th-order tensors. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Paris, pp. 911–914 (2008)
16. Barmpoutis, A., Zhuo, J.: Diffusion kurtosis imaging: robust estimation from DW-MRI using homogeneous polynomials. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Chicago, pp. 262–265 (2011)
17. Barnett, A.: Theory of Q-ball imaging redux: implications for fiber tracking. Magn. Reson. Med. **62**(4), 910–923 (2009)
18. Basser, P.J., Mattiello, J., Le Bihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson. B **103**, 247–254 (1994)
19. Basser, P.J., Pajevic, S.: A normal distribution for tensor-valued random variables: applications to diffusion tensor MRI. IEEE Trans. Med. Imaging **22**, 785–795 (2003)
20. Basser, P.J., Pajevic, S.: Spectral decomposition of a 4th-order covariance tensor: applications to diffusion tensor MRI. Signal Process. **87**, 220–236 (2007)
21. Basser, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J. Magn. Reson. Ser. B **111**, 209–219 (1996)
22. Behrens, T.E.J., Johansen-Berg, H., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? NeuroImage **34**, 144–155 (2007)
23. Bloy, L., Verma, R.: On computing the underlying fiber directions from the diffusion orientation distribution function. In: Metaxas, D.N., Axel, L., Fichtinger, G., Székely, G. (eds.) Medical Image Computing and Computer-Assisted Intervention (MICCAI), New York. LNCS, vol. 5241, pp. 1–8. Springer (2008)
24. Boothby, W.: An Introduction to Differentiable Manifolds and Riemannian Geometry. Pure and Applied Mathematics, vol. 120, 2nd edn. Academic, Orlando (1986)
25. Callaghan, P.T., Eccles, C.D., Xia, Y.: NMR microscopy of dynamic displacements: k-space and q-space imaging. J. Phys. E **21**(8), 820–822 (1988)
26. Comon, P., Golub, G., Lim, L.H., Mourrain, B.: Symmetric tensors and symmetric tensor rank. SIAM J. Matrix Anal. Appl. **30**(3), 1254–1279 (2008)
27. Correia, M.M., Newcombe, V.F., Williams, G.B.: Contrast-to-noise ratios for indices of anisotropy obtained from diffusion MRI: a study with standard clinical b-values at 3T. NeuroImage **57**(3), 1103–1115 (2011)

28. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Apparent diffusion coefficients from high angular resolution diffusion imaging: estimation and applications. Magn. Reson. Med. **56**, 395–410 (2006)
29. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical Q-Ball imaging. Magn. Reson. Med. **58**, 497–510 (2007)
30. Descoteaux, M., Deriche, R., Knösche, T.R., Anwander, A.: Deterministic and probabilistic tractography based on complex fibre orientation distributions. IEEE Trans. Med. Imaging **28**(2), 269–286 (2009)
31. De Silva, V., Lim, L.H.: Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM J. Matrix Anal. Appl. **30**(3), 1084–1127 (2008)
32. Ellingson, B.M., Cloughesy, T.F., Lai, A., Nghiemphu, P.L., Liau, L.M., Pope, W.B.: High order diffusion tensor imaging in human glioblastoma. Acad. Radiol. **18**(8), 947–954 (2011)
33. Essen, D.V., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Penna, S.D., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S., Prior, F., Schlaggar, B., Smith, S., Snyder, A., Xu, J., Yacoub, E.: The human connectome project: a data acquisition perspective. NeuroImage **62**(4), 2222–2231 (2012)
34. Florack, L., Balmashnova, E., Astola, L., Brunenberg, E.: A new tensorial framework for single-shell high angular resolution diffusion imaging. J. Math. Imaging Vis. **38**, 171–181 (2010)
35. Fuster, A., Astola, L., Florack, L.: A Riemannian scalar measure for diffusion tensor images. In: Jiang, X., Petkov, N. (eds.) Computer Analysis of Images and Patterns. LNCS, vol. 5702, pp. 419–426. Springer, Berlin/New York (2009)
36. Fuster, A., van de Sande, J., Astola, L., Poupon, C., Velterop, J., ter Haar Romeny, B.M.: Fourth-order tensor invariants in high angular resolution diffusion imaging. In: Zhang, G.H., Adluru, N. (eds.) Proceedings of the MICCAI Workshop on Computational Diffusion MRI, Toronto, pp. 54–63 (2011)
37. Gelfand, I., Kapranov, M., Zelevinsky, A.: Discriminants, Resultants, and Multidimensional Determinants. Birkhäuser, Boston (1994)
38. Geroch, R.: Mathematical Physics. Chicago Lectures in Physics. University of Chicago Press, Chicago (1985)
39. Ghosh, A., Deriche, R.: From second to higher order tensors in diffusion-MRI. In: Aja-Fernández, S., de Luis García, R., Tao, D., Li, X. (eds.) Tensors in Image Processing and Computer Vision, pp. 315–334. Springer, London (2009)
40. Ghosh, A., Deriche, R.: Fast and closed-form ensemble-average-propagator approximation from the 4th-order diffusion tensor. In: Proc. IEEE Int'l Symposium on Biomedical Imaging, pp. 1105–1108 (2010)
41. Ghosh, A., Deriche, R.: Extracting geometrical features & peak fractional anisotropy from the ODF for white matter characterization. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Chicago, pp. 266–271 (2011)
42. Ghosh, A., Deriche, R.: Generalized invariants of a 4th order tensor: building blocks for new biomarkers in dMRI. In: Panagiotaki, E., O'Donnell, L., Schultz, T., Zhang, G.H. (eds.) Proceedings of the Computational Diffusion MRI (CDMRI), Nice, pp. 165–173 (2012)
43. Ghosh, A., Deriche, R., Moakher, M.: Ternary quartic approach for positive 4th order diffusion tensors revisited. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Boston, pp. 618–621 (2009)
44. Ghosh, A., Descoteaux, M., Deriche, R.: Riemannian framework for estimating symmetric positive definite 4th order diffusion tensors. In: Metaxas, D. (ed.) MICCAI, Part I, New York, LNCS, vol. 5241, pp. 858–865 (2008)
45. Ghosh, A., Özarslan, E., Deriche, R.: Challenges in reconstructing the propagator via a cumulant expansion of the one-dimensional qspace MR signal. In: Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM), Stockholm (2010)

46. Ghosh, A., Tsigaridas, E., Mourrain, B., Deriche, R.: A polynomial approach for extracting the extrema of a spherical function and its application in diffusion MRI. Med. Image Anal. (2013, in press). doi:10.1016/j.media.2013.03.004
47. Ghosh, A., Wassermann, D., Deriche, R.: A polynomial approach for maxima extraction and its application to tractography in HARDI. In: Székely, G., Hahn, H.K. (eds.) IPMI, Kloster Irsee. LNCS, vol. 6801, pp. 723–734 (2011)
48. Grigis, A., Renard, F., Noblet, V., Heinrich, C., Heitz, F., Armspach, J.P.: A new high order tensor decomposition: application to reorientation. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Chicago, pp. 258–261 (2011)
49. Gur, Y., Jiao, F., Zhu, S.X., Johnson, C.R.: White matter structure assessment from reduced HARDI data using low-rank polynomial approximations. In: Panagiotaki, E., O'Donnell, L., Schultz, T., Zhang, G.H. (eds.) Proceedings of the Computational Diffusion MRI (CDMRI), Nice, pp. 186–197 (2012)
50. Hess, C.P., Mukherjee, P., Han, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. Magn. Reson. Med. **56**, 104–117 (2006)
51. Hillar, C., Lim, L.H.: Most tensor problems are NP-hard. JACM **60**(6), Article No. 45 (2012). Preprint, arXiv:0911.1393v2 x
52. Hitchcock, F.L.: The expression of a tensor or a polyadic as a sum of products. J. Math. Phys. **6**(1), 164–189 (1927)
53. Hitchcock, F.L.: Multiple invariants and generalized rank of a p-way matrix or tensor. J. Math. Phys. **7**(1), 39–79 (1927)
54. Hlawitschka, M., Scheuermann, G.: HOT-lines: tracking lines in higher order tensor fields. In: Silva, C., Gröller, E., Rushmeier, H. (eds.) Proceedings of the IEEE Visualization, Minneapolis, pp. 27–34 (2005)
55. Hlawitschka, M., Scheuermann, G., Anwander, A., Knösche, T., Tittgemeyer, M., Hamann, B.: Tensor lines in tensor fields of arbitrary order. In: Bebis, G., et al. (eds.) Advances in Visual Computing. LNCS, vol. 4841, pp. 341–350. Springer, Berlin/New York (2007)
56. Hui, E.S., Cheung, M.M., Qi, L., Wu, E.X.: Towards better MR characterization of neural tissues using directional diffusion kurtosis analysis. NeuroImage **42**, 122–134 (2008)
57. Hungerford, T.: Algebra. Graduate Texts in Mathematics, vol. 73. Springer, New York (1980)
58. Jansons, K.M., Alexander, D.C.: Persistent angular structure: new insights from diffusion magnetic resonance imaging data. Inverse Probl. **19**, 1031–1046 (2003)
59. Jayachandra, M.R., Rehbein, N., Herweh, C., Heiland, S.: Fiber tracking of human brain using fourth-order tensor and high angular resolution diffusion imaging. Magn. Reson. Med. **60**(5), 1207–1217 (2008)
60. Jensen, J.H., Helpern, J.A.: MRI quantification of non-Gaussian water diffusion by kurtosis analysis. NMR Biomed. **23**(7), 698–710 (2010)
61. Jensen, J.H., Helpern, J.A., Ramani, A., Lu, H., Kaczynski, K.: Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. Magn. Reson. Med. **53**, 1432–1440 (2005)
62. Jiao, F., Gur, Y., Johnson, C.R., Joshi, S.: Detection of crossing white matter fibers with high-order tensors and rank-k decompositions. In: Székely, G., Hahn, H.K. (eds.) IPMI, Kloster Irsee. LNCS, vol. 6801, pp. 538–549 (2011)
63. Kindlmann, G., Ennis, D., Whitaker, R., Westin, C.F.: Diffusion tensor analysis with invariant gradients and rotation tangents. IEEE Trans. Med. Imaging **26**(11), 1483–1499 (2007)
64. Kroonenberg, P.: Applied Multiway Data Analysis. Wiley, Hoboken (2008)
65. Kuder, T.A., Stieltjes, B., Bachert, P., Semmler, W., Laun, F.B.: Advanced fit of the diffusion kurtosis tensor by directional weighting and regularization. Magn. Reson. Med. **67**(5), 1401–1411 (2012)
66. Landsberg, J.M.: Tensors: Geometry and Applications. Graduate Studies in Mathematics, vol. 128. American Mathematical Society, Providence (2012)
67. Lang, S.: Differential and Riemannian Manifolds. Graduate Texts in Mathematics, vol. 160, 3rd edn. Springer, New York (1995)

68. Lang, S.: Algebra. Graduate Texts in Mathematics, vol. 211, rev. 3rd edn. Springer, New York (2002)
69. Lazar, M., Jensen, J.H., Xuan, L., Helpern, J.A.: Estimation of the orientation distribution function from diffusional kurtosis imaging. Magn. Reson. Med. **60**, 774–781 (2008)
70. Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M.: MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology **161**(2), 401–407 (1986)
71. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing. J. Math. Imaging Vis. **25**, 423–444 (2006)
72. Lim, L.H.: Singular values and eigenvalues of tensors: a variational approach. In: Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Puerto Vallarta, pp. 129–132 (2005)
73. Lim, L.H.: Tensors and hypermatrices. In: Hogben, L. (ed.) Handbook of Linear Algebra, 2nd edn. CRC, Boca Raton (2013)
74. Lim, L.H., Comon, P.: Nonnegative approximations of nonnegative tensors. J. Chemom. **23**(7–8), 432–441 (2009)
75. Lim, L.H., Comon, P.: Multisensor array processing: tensor decomposition meets compressed sensing. C. R. Acad. Sci. Paris **338**(6), 311–320 (2010)
76. Lim, L.H., Schultz, T.: Moment tensors and high angular resolution diffusion imaging (2013). Preprint
77. Liu, C., Bammer, R., Acar, B., Moseley, M.E.: Characterizing non-Gaussian diffusion by using generalized diffusion tensors. Magn. Reson. Med. **51**(5), 924–937 (2004)
78. Liu, C., Bammer, R., Moseley, M.E.: Generalized diffusion tensor imaging (GDTI): a method for characterizing and imaging diffusion anisotropy caused by non-Gaussian diffusion. Isr. J. Chem. **43**(1–2), 145–154 (2003)
79. Liu, C., Bammer, R., Moseley, M.E.: Limitations of apparent diffusion coefficient-based models in characterizing non-Gaussian diffusion. Magn. Reson. Med. **54**, 419–428 (2005)
80. Liu, C., Mang, S.C., Moseley, M.E.: In vivo generalized diffusion tensor imaging (GDTI) using higher-order tensors (HOT). Magn. Reson. Med. **63**, 243–252 (2010)
81. Liu, Y., Chen, L., Yu, Y.: Diffusion kurtosis imaging based on adaptive spherical integral. IEEE Signal Process. Lett. **18**(4), 243–246 (2011)
82. Lu, H., Jensen, J.H., Ramani, A., Helpern, J.A.: Three-dimensional characterization of non-Gaussian water diffusion in humans using diffusion kurtosis imaging. NMR Biomed. **19**, 236–247 (2006)
83. Minati, L., Aquino, D., Rampoldi, S., Papa, S., Grisoli, M., Bruzzone, M.G., Maccagnano, E.: Biexponential and diffusional kurtosis imaging, and generalised diffusion-tensor imaging (GDTI) with rank-4 tensors: a study in a group of healthy subjects. Magn. Reson. Mater. Phys. Biol. Med. **20**, 241–253 (2007)
84. Minati, L., Banasik, T., Brzezinski, J., Mandelli, M.L., Bizzi, A., Bruzzone, M.G., Konopka, M., Jasinski, A.: Elevating tensor rank increases anisotropy in brain areas associated with intra-voxel orientational heterogeneity (IVOH): a generalised DTI (GDTI) study. NMR Biomed. **21**(1), 2–14 (2008)
85. Mørup, M., Hansen, L., Arnfred, S., Lim, L.H., Madsen, K.: Shift invariant multilinear decomposition of neuroimaging data. NeuroImage **42**(4), 1439–1450 (2008)
86. Özarslan, E., Mareci, T.: Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. Magn. Reson. Med. **50**, 955–965 (2003)
87. Özarslan, E., Shepherd, T.M., Vemuri, B.C., Blackband, S.J., Mareci, T.H.: Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). NeuroImage **31**, 1086–1103 (2006)
88. Özarslan, E., Vemuri, B.C., Mareci, T.H.: Fiber orientation mapping using generalized diffusion tensor imaging. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Arlington, pp. 1036–1039 (2004)

89. Özarslan, E., Vemuri, B.C., Mareci, T.H.: Generalized scalar measures for diffusion MRI using trace, variance, and entropy. Magn. Reson. Med. **53**, 866–876 (2005)
90. Özarslan, E., Vemuri, B.C., Mareci, T.H.: Higher rank tensors in diffusion MRI. In: Weickert, J., Hagen, H. (eds.) Visualization and Processing of Tensor Fields, chap. 10, pp. 177–187. Springer, Berlin (2006)
91. Poot, D.H.J., den Dekker, A.J., Achten, E., Verhoye, M., Sijbers, J.: Optimal experimental design for diffusion kurtosis imaging. IEEE Trans. Med. Imaging **29**(3), 819–829 (2010)
92. Qi, L.: Eigenvalues of a real supersymmetric tensor. J. Symb. Comput. **40**, 1302–1324 (2005)
93. Qi, L., Han, D., Wu, E.X.: Principal invariants and inherent parameters of diffusion kurtosis tensors. J. Math. Anal. Appl. **349**, 165–180 (2009)
94. Qi, L., Wang, Y., Wu, E.X.: D-eigenvalues of diffusion kurtosis tensors. J. Comput. Appl. Math. **221**, 150–157 (2008)
95. Qi, L., Yu, G., Wu, E.X.: Higher order positive semidefinite diffusion tensor imaging. SIAM J. Imaging Sci. **3**(3), 416–433 (2010)
96. Renard, F., Noblet, V., Heinrich, C., Kremer, S.: Reorientation strategies for high order tensors. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Rotterdam, pp. 1185–1188 (2010)
97. Schultz, T.: Learning a reliable estimate of the number of fiber directions in diffusion MRI. In: Ayache, N., et al. (eds.) Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI) Part III, Nice. LNCS, vol. 7512, pp. 493–500 (2012)
98. Schultz, T., Kindlmann, G.: A maximum enhancing higher-order tensor glyph. Comput. Graph. Forum **29**(3), 1143–1152 (2010)
99. Schultz, T., Schlaffke, L., Schölkopf, B., Schmidt-Wilcke, T.: HiFiVE: a hilbert space embedding of fiber variability estimates for uncertainty modeling and visualization. Comput. Graph. Forum **32**(3), 121–130 (2013)
100. Schultz, T., Seidel, H.P.: Estimating crossing fibers: a tensor decomposition approach. IEEE Trans. Vis. Comput. Graph. **14**(6), 1635–1642 (2008)
101. Schultz, T., Westin, C.F., Kindlmann, G.: Multi-diffusion-tensor fitting via spherical deconvolution: a unifying framework. In: Jiang, T., et al. (eds.) Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI), Beijing. LNCS, vol. 6361, pp. 673–680. Springer (2010)
102. Sidiropoulos, N., Bro, R., Giannakis, G.: Parallel factor analysis in sensor array processing. IEEE Trans. Signal Process. **48**(8), 2377–2388 (2000)
103. Smilde, A., Bro, R., Geladi, P.: Multi-way Analysis: Applications in the Chemical Sciences. Wiley, West Sussex (2004)
104. Struik, D.J.: Schouten, Levi-Civita and the emergence of tensor calculus. In: Rowe, D., McCleary, J. (eds.) History of Modern Mathematics, vol. 2, pp. 99–105. Academic, Boston (1989)
105. Tabesh, A., Jensen, J.H., Ardekani, B.A., Helpern, J.A.: Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. Magn. Reson. Med. **65**, 823–836 (2011)
106. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. NeuroImage **35**, 1459–1472 (2007)
107. Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage **23**, 1176–1185 (2004)
108. Tristan-Vega, A., Westin, C.F., Aja-Fernandez, S.: A new methodology for the estimation of fiber populations in the white matter of the brain with the funk-radon transform. NeuroImage **49**(2), 1301–1315 (2010)
109. Tuch, D.S.: Q-ball imaging. Magn. Reson. Med. **52**, 1358–1372 (2004)
110. Vasilescu, M., Terzopoulos, D.: Multilinear image analysis for facial recognition. Proc. Int. Conf. Pattern Recognit (ICPR) **2**, 511–514 (2002)
111. Veraart, J., Van Hecke, W., Sijbers, J.: Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a Rician noise model. Magn. Reson. Med. **66**, 678–686 (2011)

112. Warner, F.: Foundations of Differentiable Manifolds and Lie Groups. Graduate Texts in Mathematics, vol. 94. Springer, New York/Berlin (1983)
113. Weldeselassie, Y.T., Barmpoutis, A., Atkins, M.S.: Symmetric positive-definite Cartesian tensor orientation distribution functions (CT-ODF). In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI), Beijing. LNCS, vol. 6361, pp. 582–589 (2010)
114. Westin, C.F., Peled, S., Gudbjartsson, H., Kikinis, R., Jolesz, F.A.: Geometrical diffusion measures for MRI from tensor basis analysis. In: Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM), Vancouver, p. 1742 (1997)
115. Ying, L., Zou, Y.M., Klemer, D.P., Wang, J.J.: Determination of fiber orientation in MRI diffusion tensor imaging based on higher-order tensor decomposition. In: Proceedings of the International Conference on IEEE Engineering in Medicine and Biology Society (EMBS), Lyon, pp. 2065–2068 (2007)
116. Yokonuma, T.: Tensor Spaces and Exterior Algebra. Translations of Mathematical Monographs, vol. 108. American Mathematical Society, Providence (1992)

Fourth Order Symmetric Tensors and Positive ADC Modeling

Aurobrata Ghosh and Rachid Deriche

Abstract High Order Cartesian Tensors (HOTs) were introduced in Generalized DTI (GDTI) to overcome the limitations of DTI. HOTs can model the apparent diffusion coefficient (ADC) with greater accuracy than DTI in regions with fiber heterogeneity. Although GDTI HOTs were designed to model positive diffusion, the straightforward least square (LS) estimation of HOTs doesn't guarantee positivity. In this chapter we address the problem of estimating 4th order tensors with positive diffusion profiles.

Two known methods exist that broach this problem, namely a Riemannian approach based on the algebra of 4th order tensors, and a polynomial approach based on Hilbert's theorem on non-negative ternary quartics. In this chapter, we review the technicalities of these two approaches, compare them theoretically to show their pros and cons, and compare them against the Euclidean LS estimation on synthetic, phantom and real data to motivate the relevance of the positive diffusion profile constraint.

1 Introduction

Diffusion Tensor Imaging (DTI) [1, 2] has become the de facto standard today in diffusion MRI (dMRI) for investigating the complex microstructure of the cerebral white matter in-vivo and non-invasively. Its tremendous popularity is due to its simplicity in acquisition requisites and elegance in interpretation, which makes it easy to implement the technique and infer the white matter microstructure, in particular the underlying fiber orientations. Based on Fick's phenomenological anisotropic diffusion equation, the DTI signal for the diffusion gradient $\mathbf{G}$, is

A. Ghosh (✉) • R. Deriche
INRIA Sophia Antipolis-Méditerranée, Sophia Antipolis, France
e-mail: aurobrata.ghosh@gmail.com; rachid.deriche@inria.fr

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_7,

described by the modified Stejskal-Tanner equation parameterized by the second order diffusion tensor **D** [3]:

$$S = S_0 \exp\left(-b\mathbf{g}^T\mathbf{D}\mathbf{g}\right), \tag{1}$$

where $b = \gamma^2\delta^2 g^2\left(\Delta - \frac{\delta}{3}\right)$, $g = |\mathbf{G}|$, and $\mathbf{g} = \mathbf{G}/|\mathbf{G}|$. In DTI, the apparent diffusion coefficient (ADC) is modeled by the spherical function $D(\mathbf{g}) = \mathbf{g}^T\mathbf{D}\mathbf{g}$. However, in spite of its usefulness, it is well known that DTI is inherently limited in regions with heterogeneous fiber distributions, such as in fiber-crossings. In such regions DTI can neither accurately model the complex shape of the resulting ADC, nor correctly infer the underlying fiber bundle layout.

Generalized DTI (GDTI) [4], was proposed to overcome this limitation by modeling the complex shaped ADC with greater accuracy using Cartesian tensors of order higher than two, the so called higher order (diffusion) tensors (HOTs). GDTI, like DTI, is also based on Fick's phenomenological laws of diffusion, where the diffusion tensor is replaced by a spherical diffusion function parameterized by a HOT, or as its projection on to the unit sphere. The GDTI signal for the diffusion gradient **G** is similarly described by:

$$S = S_0 \exp\left(-bD(\mathbf{g})\right), \quad D(\mathbf{g}) = \sum_{j_1=1}^{3}\sum_{j_2=1}^{3}\cdots\sum_{j_k=1}^{3} D_{j_1,j_2\ldots j_k} g_{j_1} g_{j_2}\ldots g_{j_k}, \tag{2}$$

where, $D_{j_1,j_2\ldots j_k}$ are the coefficients of the kth order, three dimensional, diffusion HOT $\mathscr{D}^{(k)}$, and g_{j_i} are the components of the unit gradient vector **g**. The complex shaped ADC is described in GDTI by $D(\mathbf{g})$. Since **g** is a unit norm vector, it can also be described by the two parameters $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$ as $\mathbf{g} = [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]^T = [g_x, g_y, g_z]^T$, which shows that the ADC or the spherical diffusion function is the projection of $\mathscr{D}^{(k)}$ on to the unit sphere.

This form of the diffusion function helps derive certain properties of the diffusion HOT which greatly simplifies the GDTI model [4]. First, when k is odd $D(-\mathbf{g}) = -D(\mathbf{g})$. However, since negative diffusion is non-physical, this implies that k can only be even, or only even ordered HOTs are of interest in modeling the ADC. Second, although a kth order 3D HOT can have 3^k independent coefficients, since only its projection along a vector **g** is of interest – $\mathscr{D}^{(k)}$ has to be symmetric – or its coefficients should be equal under any permutation σ, of the coefficient indices $D_{j_1,j_2\ldots j_k} = D_{\sigma(j_1,j_2\ldots j_k)}$. This reduces the number of independent coefficients of the kth order HOT to a more tractable:

$$N_k = \frac{(k+1)(k+2)}{2}. \tag{3}$$

In other words, to describe the ADC more accurately using GDTI, it is required to estimate from the diffusion signal the coefficients of a 3D symmetric HOT of even rank, such that the diffusion function or the estimated ADC is positive.

The independent coefficients of the kth order diffusion HOT are in practice estimated using the least squares (LS) approach [4] in a fashion almost identical to the approach for estimating the six coefficients of the diffusion tensor in DTI. The LS approach, although, rapid, since it involves only linear operations, does not guarantee that the estimated HOT will result in a positive diffusion function even when k is considered even. In other words, the reason for considering k to be even, i.e. the estimated ADC should be positive, is not satisfied by the LS estimation.

In this chapter we present two approaches for estimating, in particular 4th order, diffusion HOTs from the diffusion signal that guarantee that the estimated ADC or the diffusion function is positive. In the first method, we take recourse to the fact that 3D symmetric 4th order tensors can be rewritten through a mapping as 6D symmetric 2nd order tensors. This makes it possible to reformulate the problem of estimating a 4th order tensor with a positive diffusion profile, to a problem of estimating a 2nd order tensor with a positive diffusion profile, albeit in 6D. We solve this problem by applying the Riemannian framework developed for symmetric positive definite (SPD) tensors of order 2, for estimating DTI diffusion tensors with positive diffusion profiles.

In the second method, we base ourselves on the polynomial interpretation of HOTs. Therefore, the diffusion function $D(\mathbf{g})$ is re-interpreted as a homogeneous polynomial in the components of the unit norm gradient vector $\mathbf{g}$. This allows for a powerful parameterization of the diffusion signal, which ensures that the estimation process guarantees a 4th order HOT with a positive diffusion profile. This parameterization comes from the properties of ternary quartics, which was first pointed out in [5, 6]. Also it has been proposed in [7] that the affine invariant Riemannian metric may not be well suited for diffusion data. The polynomial parameterization, therefore, provides an alternative approach for estimating 4th order diffusion tensors with positive diffusion profiles, which employs the Euclidean metric that is better suited for handling diffusion data [7].

We note that solutions to the problem of estimating arbitrary even ordered HOTs with the positivity constraint have also been proposed in [20] and [8]. These methods and the contents of this chapter can be seen briefly resumed in chapter "Higher-Order Tensors in Diffusion Imaging". However, in this chapter we present in greater detail the particular problem of estimating 4th order tensors with the positivity constraint, since 4th order tensors commonly appear in many problems, such as Diffusion Kurtosis Imaging (DKI: see again chapter "Higher-Order Tensors in Diffusion Imaging"). The importance of the methods presented here is highlighted by the fact that these methods have been recently used to estimate 4th order kurtosis tensors with positivity constraint [9].

This chapter is structured as follows. Section 2 is devoted to the Riemannian approach. Sections 2.1 and 2.2 present the algebra of 2nd and 4th order tensors which allow us to formulate the Riemannian framework. The Riemannian estimation scheme is put together in Sect. 2.3. Section 3 is devoted to the ternary quartic approach, with first the theory and then the algorithm in Sect. 3.3. Experiments and results are described and discussed in Sect. 4. We conclude in Sect. 5.

2 A Riemannian Approach for Symmetric Positive Definite Fourth Order Diffusion Tensors

The problem of estimating a diffusion tensor from the signal, which satisfies the positive diffusion profile has been extensively considered in DTI. Negative diffusion, which is non-physical, can also be a problem while estimating a 2nd order diffusion tensor $\mathbf{D}$, which happens when the DTI-ADC $\mathbf{g}^T\mathbf{D}\mathbf{g} < 0$, for some gradient direction $\mathbf{g}$. This can occur since the LS estimation process doesn't guarantee that the diffusion tensor will have a positive diffusion profile. This condition requires a dedicated mathematical framework which constraints the estimation process to only diffusion tensors $\mathbf{D}$ such that $\mathbf{g}^T\mathbf{D}\mathbf{g} > 0,\ \forall \mathbf{g} \in S^2$.

An adequate framework for such an estimation was proposed by identifying the appropriate set of 2nd order tensors that satisfy the positive quadratic form, namely $\mathcal{S}ym_n^+$, the set of SPD matrices, which satisfy $\mathbf{x}^T\Sigma\mathbf{x} > 0,\ \forall \mathbf{x} \in \mathbf{R}^n\backslash\{\mathbf{0}\}$, and $\Sigma \in \mathcal{S}ym_n^+$. In other words, if the estimation process were to only operate in the space of $\mathcal{S}ym_3^+$ (in the case of DTI, $n = 3$), then the estimated diffusion tensor would satisfy the positive diffusion profile. The mathematical framework that was proposed, which allows to do this consists of an affine invariant metric of $\mathcal{S}ym_n^+$, the Riemannian metric [10–13], and a similarity invariant metric of $\mathcal{S}ym_n^+$, the Log-Euclidean metric [14], which naturally confine operations on SPD matrices to the space of $\mathcal{S}ym_n^+$.

Deriving an equivalent Riemannian metric for the space of 4th order diffusion tensors would, however, be far more involved due to the increase in order or the multi-linear property of HOTs. Nonetheless, such a metric would be the right framework to use in the estimation process of the 4th order diffusion tensor, since it would ensure that the estimated HOT satisfies the positive diffusion profile. However, given the symmetry condition of a diffusion HOT, this problem can be simplified by reformulating the diffusion profile of a 4th order HOT (Eq. (2)) to a bilinear form dependent on a 2nd order tensor. Mathematically, this would convert the problem to the case of estimating a 2nd order tensor in $\mathcal{S}ym_n^+$, like in DTI. However, the conversion from a symmetric 4th order 3D tensor, results in a symmetric 2nd order tensor in 6D [15–17]. Therefore, we would have to consider the space of $\mathcal{S}ym_6^+$ instead of the space of $\mathcal{S}ym_3^+$.

In this section, we propose to use this approach of transforming a symmetric 3D 4th order Cartesian diffusion tensor to a symmetric 6D 2nd order tensor, and of applying the Riemannian metric of the space $\mathcal{S}ym_6^+$, to estimate a 4th order diffusion tensor from the signal with a positive diffusion profile in GDTI [18].

2.1 Algebra of Second Order Tensors

To understand the algebra of 4th order tensors, which is required to manipulate these entities, and to transform them to isometrically equivalent 2nd order tensors,

we start with 2nd order tensors, which are well studied and intuitively easy to understand. Much of the following formulation of Cartesian 2nd and 4th order tensors in an Euclidean space can be found in [15, 16], where, essentially a tensor is used interchangeably with the matrix of a linear transformation.

Given an n dimensional inner product space (vector space with an inner product) V, an nD 2nd order tensor $\mathbf{A} = \mathscr{A}^{(2)}$ is defined as the $n \times n$ matrix of the linear transformation:

$$A : V \to V, \quad \text{st} \quad \mathbf{x} \to \mathbf{A}\mathbf{x}, \ \mathbf{x} \in V. \tag{4}$$

The transpose of the linear transformation, with matrix $\mathbf{A}^T$, can be defined from the inner product of V as $\langle \mathbf{x}, \mathbf{A}^T\mathbf{y} \rangle = \langle \mathbf{A}\mathbf{x}, \mathbf{y} \rangle , \quad \forall \mathbf{x}, \mathbf{y} \in V$. The space of linear transformations from V to V, itself forms a vector space, which can be called $\text{Lin}(V) = \{A : V \to V\}$. The transpose of A can be used to define a natural inner product on $\text{Lin}(V)$ (summation over repeated indices over their whole range):

$$\langle \mathbf{A}, \mathbf{B} \rangle := tr(\mathbf{A}^T\mathbf{B}) = A_{ij}B_{ij}, \quad A, B \in \text{Lin}(V). \tag{5}$$

If V is $\mathbf{R}^n$, then $\text{Lin}(V)$ is $\mathbf{R}^{n\times n}$, and it is isomorphic to $\mathbf{R}^{n^2}$. Therefore a tensor $\mathbf{A}$ in $\mathbf{R}^{n\times n}$ can be written as a vector $\mathbf{a}$, in $\mathbf{R}^{n^2}$. Furthermore, the isomorphism is an isometry, since $\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{A}, \mathbf{B} \rangle$, where the first inner product is the natural inner product of the vector space $\mathbf{R}^{n^2}$, and the second inner product is the newly defined inner product of $\text{Lin}(V) = \mathbf{R}^{n\times n}$.

A symmetric linear transformation A from V to V, can be defined from the transpose of its corresponding 2nd order tensor, as $\mathbf{A} = \mathbf{A}^T$, which in terms of its components can be described by $A_{ij} = A_{ji}$. It is then possible to decompose a 2nd order tensor (or linear transformation) into its symmetric and skew-symmetric parts by $\mathbf{A}^s = (\mathbf{A} + \mathbf{A}^T)/2$ and $\mathbf{A}^a = (\mathbf{A} - \mathbf{A}^T)/2$ respectively, such that $\mathbf{A} = \mathbf{A}^s + \mathbf{A}^a$.

Finally the space of symmetric linear transformations $\text{Sym}(V) = \{A \in \text{Lin}(V) | \mathbf{A} = \mathbf{A}^T\}$, forms a subspace of $\text{Lin}(V)$. Since, an nD symmetric 2nd order tensor has $n(n+1)/2$ independent coefficients, if V is $\mathbf{R}^n$, then $\text{Sym}(V)$ is isomorphic to $\mathbf{R}^{n(n+1)/2}$, and this mapping can be established in such a fashion that it is also an isometry, just like in the case of $\text{Lin}(V)$, or $\langle \mathbf{a}_s, \mathbf{b}_s \rangle = \langle \mathbf{A}^s, \mathbf{B}^s \rangle$, for $\mathbf{a}_s, \mathbf{b}_s \in \mathbf{R}^{n(n+1)/2}$ and $A^s, B^s \in \text{Sym}(V)$. An example for such an isometric mapping when $n = 3$, can be established between a symmetric 3D 2nd order tensor $\mathbf{B}$, and $\mathbf{b}$, a vector or a 6D 1st order tensor:

$$\mathbf{b} = [B_{11}, B_{22}, B_{33}, \sqrt{2}B_{12}, \sqrt{2}B_{13}, \sqrt{2}B_{23}]^T, \tag{6}$$

where B_{ij} are the coefficients of $\mathbf{B}$.

2.2 Algebra of Fourth Order Tensors

The background for understanding the algebra of 4th order tensors is formed by the definition of the inner product, the isometric mapping to vectors (1st order tensors) of higher dimensions, and the symmetry properties, in particular $\mathrm{Sym}(V)$, of the space of 2nd order tensors or $\mathrm{Lin}(V)$. In an analogous way, we will define 4th order tensors as linear transformations from a vector space onto itself, define an inner product for the vector space of these linear transformations, study their symmetries, and establish an isometric mapping from the linear transformations to a vector space of lower order and higher dimension, which will allow us to manipulate 4th order tensors as 2nd order tensors.

The algebra of 4th order tensors can be described by proceeding in exactly the way as done above for 2nd order tensors, but with $\mathrm{Lin}(V)$ as the vector space in place of V. Let an nD 4th order tensor $\hat{\mathscr{A}} = \mathscr{A}^{(4)}$ be defined as the $n \times n \times n \times n$ transformation array of the linear transformation (summation over repeated indices over their whole range):

$$\mathrm{A} : \mathrm{Lin}(V) \to \mathrm{Lin}(V), \quad \text{st} \quad \mathbf{C} \to \hat{\mathscr{A}}\mathbf{C} = A_{ijkl}C_{kl}, \ \mathbf{C} \in \mathrm{Lin}(V). \tag{7}$$

Since an inner product for $\mathrm{Lin}(V)$ exists, it can be used to define the transpose of the linear transformation, with the transformation array $\hat{\mathscr{A}}^T$, as:

$$\left\langle \mathbf{D}, \hat{\mathscr{A}}^T\mathbf{C} \right\rangle = \left\langle \hat{\mathscr{A}}\,\mathbf{D}, \mathbf{C} \right\rangle, \quad \forall \mathbf{C}, \mathbf{D} \in \mathrm{Lin}(V). \tag{8}$$

Again the space of linear transformations from $\mathrm{Lin}(V)$ to $\mathrm{Lin}(V)$ forms a vector space, which can be called $\mathscr{L}in(V) = \{\mathrm{A} : \mathrm{Lin}(V) \to \mathrm{Lin}(V)\}$, and again the transpose of A can be used to define an inner product on $\mathscr{L}in(V)$ (summation over repeated indices over their whole range):

$$\left\langle \hat{\mathscr{A}}, \hat{\mathscr{B}} \right\rangle := tr(\hat{\mathscr{A}}^T\hat{\mathscr{B}}) = A_{ijkl}B_{ijkl}, \quad \hat{\mathscr{A}}, \hat{\mathscr{B}} \in \mathscr{L}in(V). \tag{9}$$

If V is $\mathbf{R}^n$, then $\mathrm{Lin}(V)$ is $\mathbf{R}^{n\times n}$, and $\mathscr{L}in(V)$ is $\mathbf{R}^{n\times n\times n\times n}$, which is isomorphic to $\mathbf{R}^{n^4}$. Therefore an nD 4th order tensor can be written as a vector in $\mathbf{R}^{n^4}$. However, of greater interest is that $\mathscr{L}in(V)$ is also isomorphic to $\mathbf{R}^{n^2\times n^2}$, which implies that an nD 4th order tensor $\mathscr{A}$ can be written as an n^2D 2nd order tensor $\mathbf{A}$. Furthermore, this isomorphism is also an isometry $\langle \mathbf{A}, \mathbf{B} \rangle = \left\langle \hat{\mathscr{A}}, \hat{\mathscr{B}} \right\rangle$.

Symmetries of 4th order tensors present a richer set of possibilities than the symmetry of 2nd order tensors, since a number of symmetries can be defined by applying different "symmetry rules" on the four coefficient indices. Indeed, we shall present the major symmetry, the minor symmetry and the total symmetry. Total symmetry is, however, the symmetry of interest to us, which in the mathematical approach to tensors is the *definition of symmetry* of a HOT, where the coefficients of the HOT remain unchanged under any permutation of the coefficient indices. This is

also the symmetry condition required by the diffusion HOT in GDTI, as implied by its properties. However, this symmetry is best called total symmetry (or complete symmetry), to differentiate it from the other possible symmetries that are derived from physics and that carry important physical interpretations.

We shall, however, not present such physical interpretations here, but content ourselves with counting the number of independent coefficients of a 4th order tensor under the various symmetries. To do this we will require the formula for counting the number of ways of choosing m elements from n elements without order and with repetition (combination) $S_{m,n} = \binom{n+m-1}{m}$.

Major symmetry of an nD 4th order tensor $\mathscr{A}$ is defined by the index symmetry rule $A_{ij,kl} = A_{kl,ij}$. To count the number of independent coefficients of $\mathscr{A}$, which satisfies major symmetry, we consider the isometrically equivalent n^2D 2nd order tensor $\mathbf{A}$, which has only two indices $I = ij$ and $J = kl$. Therefore, major symmetry of $\mathscr{A}$ can be translated as the index symmetry rule of $\mathbf{A}$ as $\hat{A}_{IJ} = \hat{A}_{JI}$, where $\hat{A}_{o_1 o_2}$ are the coefficients of $\mathbf{A}$, which implies that $\mathbf{A} = \mathbf{A}^T$. Therefore, the number of independent coefficients of $\mathscr{A}$, which satisfies major symmetry, is ($\overline{M}$ is used to indicate major symmetry):

$$N_{\overline{M}} = \frac{n^2(n^2+1)}{2}. \tag{10}$$

Note that major symmetry for $\mathscr{A}$, corresponds to the regular notion of symmetry for the 2nd order tensor $\mathbf{A}$. Therefore, symmetry properties of $\mathbf{A}$, such as decomposition into a symmetric part and a skew symmetric part and eigen-decomposition, can be attributed to the 4th order tensor $\mathscr{A}$ by isomorphism. Major symmetry also corresponds to the notion of symmetry induced by the definition of the transpose of a 4th order tensor, or a linear transformation from $\mathrm{Lin}(V)$ to $\mathrm{Lin}(V)$.

Minor symmetry of an nD 4th order tensor $\mathscr{A}$ is defined by the index symmetry rule $A_{ij,kl} = A_{ji,kl} = A_{ij,lk}$. To count the number of independent coefficients of $\mathscr{A}$, which satisfies minor symmetry, the index rule can be seen as first choosing 2 index values $\{ij\}$ from n index values without order and with repetition, and then again choosing 2 index values $\{lk\}$ under the same condition. However, since $\{ij\}$ and $\{lk\}$ don't swap, their mutual order is important. Therefore, the number of independent coefficients of $\mathscr{A}$, which satisfies minor symmetry is ($\underline{M}$ is used to indicate minor symmetry):

$$N_{\underline{M}} = \binom{n+2-1}{2}^2 = \frac{n^2(n+1)^2}{4}. \tag{11}$$

The number of independent coefficients of an nD 4th order tensor with combined major and minor symmetries can be computed by combining the reasonings of the individual counts. First choose 2 index values $\{ij\} = I$ or $\{lk\} = J$ from n index values without order and with repetition, which gives $\sqrt{N_{\underline{M}}}$. Then choose 2

index values $\{IJ\}$ from these $\sqrt{N_{\underline{M}}}$ index values without order and with repetition. Therefore, the number of independent coefficients of $\mathscr{A}$, which satisfies both major and minor symmetries is:

$$N_{(\overline{M}+\underline{M})} = \binom{\sqrt{N_{\underline{M}}} + 2 - 1}{2}. \tag{12}$$

Total symmetry or just symmetry, is defined for an nD 4th order tensor $\mathscr{A}$ by the index symmetry rule $A_{ijkl} = A_{\sigma(ijkl)}$, where $\sigma(ijkl)$ is any permutation of the indices $\{ijkl\}$. This is the symmetry satisfied by any HOT in the GDTI model, which implies from Eq. (3), that the number of independent coefficients for a 3D kth order GDTI HOT is N_k. However, the number of independent coefficients of an nD 4th order tensor $\mathscr{A}$, which satisfies total symmetry can also be counted as the number of ways of choosing 4 index values from n possible index values, therefore:

$$N_T = \binom{n + 4 - 1}{4}. \tag{13}$$

If we consider $k = 4$, it implies $N_k = 15$, and if we consider $n = 3$, it implies $N_T = 15$. This establishes the consistency between N_k and N_T.

Any 4th order tensor $\mathscr{A}$ satisfying major and minor symmetries can be decomposed in a unique manner into a totally symmetric 4th order tensor $\mathscr{A}^s$ and its asymmetric part $\mathscr{A}^a$ such that $\mathscr{A} = \mathscr{A}^s + \mathscr{A}^a$. The coefficients of the totally symmetric part and the asymmetric part can be computed from [15]:

$$\begin{aligned} A^s_{ijkl} &= \tfrac{1}{3}\left(A_{ijkl} + A_{ikjl} + A_{ilkj}\right) \\ A^a_{ijkl} &= \tfrac{1}{3}\left(2A_{ijkl} - A_{ikjl} - A_{ilkj}\right). \end{aligned} \tag{14}$$

These, along with the definition of the inner product between two 4th order tensors can be used to show that $\langle \mathscr{A}^s, \mathscr{B}^a \rangle = tr(\mathscr{A}^s \mathscr{B}^a) = 0$.

These symmetries greatly reduce the number of independent coefficients of an nD 4th order tensor from the total number of possible independent coefficients, which is n^4. Of particular interest are the 4th order tensors which satisfy both major and minor symmetries. These form a subspace of $\mathscr{L}in(V)$, called:

$$\mathscr{S}ym_{(\overline{M}+\underline{M})}(V) = \{\mathrm{A} : \mathscr{L}in(V) \to \mathscr{L}in(V) | \mathscr{A} \text{ satisfies major \& minor symmetries}\}, \tag{15}$$

which is isometrically isomorphic to $\mathbf{R}^{N_{(\overline{M}+\underline{M})}}$.

When $n = 3$, $N_{\underline{M}} = 36$, and $N_{(\overline{M}+\underline{M})} = 21$. Therefore, $\mathscr{S}ym_{(\overline{M}+\underline{M})}(V)$ is isomorphic to $\mathbf{R}^{21}$, which is the space of symmetric 6D 2nd order tensors. An example of an isometric isomorphism that can be established in this case between a 3D 4th order tensor $\mathscr{A}_{(\overline{M}+\underline{M})}$ and a 6D 2nd order tensor $\mathbf{A}$ is [19]:

$$\mathbf{A} = \begin{pmatrix} A_{xxxx} & A_{xxyy} & A_{xxzz} & \sqrt{2}A_{xxxy} & \sqrt{2}A_{xxxz} & \sqrt{2}A_{xxyz} \\ A_{xxyy} & A_{yyyy} & A_{yyzz} & \sqrt{2}A_{yyxy} & \sqrt{2}A_{yyxz} & \sqrt{2}A_{yyyz} \\ A_{xxzz} & A_{yyzz} & A_{zzzz} & \sqrt{2}A_{zzxy} & \sqrt{2}A_{zzxz} & \sqrt{2}A_{zzyz} \\ \sqrt{2}A_{xxxy} & \sqrt{2}A_{yyxy} & \sqrt{2}A_{zzxy} & 2A_{xyxy} & 2A_{xyxz} & 2A_{xyyz} \\ \sqrt{2}A_{xxxz} & \sqrt{2}A_{yyxz} & \sqrt{2}A_{zzxz} & 2A_{xyxz} & 2A_{xzxz} & 2A_{xzyz} \\ \sqrt{2}A_{xxyz} & \sqrt{2}A_{yyyz} & \sqrt{2}A_{zzyz} & 2A_{xyyz} & 2A_{xzyz} & 2A_{yzyz} \end{pmatrix}, \tag{16}$$

where A_{ijkl} are the independent coefficients of $\mathscr{A}_{(\overline{M}+\underline{M})}$. This map, along with the map in Eq. (6), which transforms a symmetric 2nd order tensor to a vector or a 1st order tensor, allows us to isometrically rewrite the effects of a linear transformation $\mathrm{A}_{(\overline{M}+\underline{M})}$ in $\mathscr{S}ym_{(\overline{M}+\underline{M})}(V)$ on a symmetric linear transformation B^s in $\mathrm{Sym}(V)$, as a matrix vector product when $n = 3$:

$$\mathscr{A}_{(\overline{M}+\underline{M})}\mathbf{B}^s = \mathbf{A}_{(\overline{M}+\underline{M})}\mathbf{b}^s, \tag{17}$$

$$\left\langle \mathscr{D}^s, \mathscr{A}_{(\overline{M}+\underline{M})}\mathbf{B}^s \right\rangle = \mathbf{d}^{s^T}\mathbf{A}_{(\overline{M}+\underline{M})}\mathbf{b}^s. \tag{18}$$

However, since diffusion HOTs from the GDTI model have to satisfy total symmetry, we are interested in the space of 3D 4th order tensors, which satisfy total symmetry. These also form a subspace of $\mathscr{L}in(V)$, called:

$$\mathscr{S}ym_T(V) = \{\mathrm{A} : \mathscr{L}in(V) \to \mathscr{L}in(V) | \mathscr{A} \text{ satisfies total symmetry}\}, \tag{19}$$

which is isometrically isomorphic to $\mathbf{R}^{15}$, since $N_T = 15$ when $n = 3$. Although $\mathbf{R}^{15}$ corresponds to the space of symmetric 5D 2nd order tensors, the isometry to symmetric 6D 2nd order tensors (Eq. (16)) can be modified to represent $\mathscr{S}ym_T(V)$, with the added equalities:

$$\begin{aligned} A_{xxyy} = A_{xyxy};\ A_{xxzz} = A_{xzxz};\ A_{yyzz} = A_{yzyz} \\ A_{xxyz} = A_{xyxz};\ A_{yyxz} = A_{xyyz};\ A_{zzxy} = A_{xzyz}. \end{aligned} \tag{20}$$

Applying these equalities to $\mathbf{A}$ in Eq. (16), is equivalent to decomposing the 3D 4th order tensor $\mathscr{A}_{(\overline{M}+\underline{M})}$, with major and minor symmetries, into its totally symmetric part $\mathscr{A}^s_{(\overline{M}+\underline{M})}$ [15]. In other words, an isometry from $\mathscr{S}ym_T(V)$ to the space of symmetric 6D 2nd order tensors can be established by considering the totally symmetric part of the equivalent 3D 4th order tensor with only major and minor symmetries.

The final isometry between $\mathscr{S}ym_T(V)$ and the space of symmetric 6D 2nd order tensors is the transformation that converts a 3D 4th order diffusion tensor from the GDTI model to an isometrically equivalent symmetric 6D 2nd order tensor. This allows us to use the Riemannian metric on the space of $\mathscr{S}ym_6^+$, to estimate the 4th order diffusion tensor with a positive diffusion profile.

2.3 Estimating a SPD Fourth Order Diffusion Tensor

First we re-write the diffusion function in Eq. (2), which is written in terms of the coefficients of the kth order tensor $\mathscr{D}^{(k)}$ and of the unit gradient vector $\mathbf{g}$, in the tensor terminology when $k = 4$:

$$D(\mathbf{g}) = \left\langle \mathscr{D}^{(4)}, \mathscr{G} \right\rangle, \qquad \text{where } \mathscr{G} = \mathbf{g} \otimes \mathbf{g} \otimes \mathbf{g} \otimes \mathbf{g} \tag{21}$$

$$= \left\langle \mathbf{B}, \mathscr{D}^{(4)}\mathbf{B} \right\rangle, \qquad \text{where } \mathbf{B} = \mathbf{g} \otimes \mathbf{g} \tag{22}$$

$$= \left\langle \mathbf{b}, \hat{\mathbf{D}}\mathbf{b} \right\rangle = \mathbf{b}^T \hat{\mathbf{D}} \mathbf{b}, \tag{23}$$

where $\mathscr{D}^{(4)}$ is the 4th order diffusion HOT in GDTI, $\mathscr{G}$ is a totally symmetric 4th order tensor computed from the outer products "$\otimes$" of the gradient vector, similarly $\mathbf{B}$ is a symmetric 2nd order tensor computed from the outer products of $\mathbf{g}$, $\mathbf{b}$ is the vector form of $\mathbf{B}$ using the isometric map from Eq. (6), and $\hat{\mathbf{D}}$ is the symmetric 6D matrix form of $\mathscr{D}^{(4)}$ using the isometric map from Eq. (16). The first two equalities can be derived from the coefficients' equation in Eq. (2), and the third equality can be derived from Eqs. (17) to (18). Therefore, the diffusion signal from the GDTI model (Eq. (2)) when $k = 4$, can be written in tensor form as:

$$S = S_0 \exp\left(-b\mathbf{b}^T \hat{\mathbf{D}} \mathbf{b}\right). \tag{24}$$

In this form, the problem of estimating the 4th order diffusion tensor $\mathscr{D}^{(4)}$, from the signal, with a positive diffusion profile can be solved by estimating the 2nd order tensor $\hat{\mathbf{D}}$, from the signal, in $\mathscr{S}ym_6^+$.

The objective function we minimize to estimate $\hat{\mathbf{D}}$ from N diffusion weighted images (DWIs) is the linearized form of the modified GDTI Stejskal-Tanner equation:

$$E(\hat{\mathbf{D}}) = \frac{1}{2} \sum_{i=1}^{N} \left(\frac{1}{b} \ln\left(\frac{S_i}{S_0}\right) + \mathbf{b}_i^T \hat{\mathbf{D}} \mathbf{b}_i \right)^2. \tag{25}$$

To estimate $\hat{\mathbf{D}}$ in $\mathscr{S}ym_6^+$, we have to consider the Riemannian manifold of $\mathscr{S}ym_6^+$, and the appropriate gradient descent in that manifold. These can be derived from the details of the Riemannian framework presented in [10–13]. It requires computing the gradient of $E(\hat{\mathbf{D}})$ in that manifold, which at every point in $\mathscr{S}ym_6^+$ is defined from the directional derivatives in the corresponding tangent plane.

The Riemannian gradient of $E(\hat{\mathbf{D}})$ at $\hat{\mathbf{D}}$ in the manifold $\mathscr{S}ym_6^+$ is [13]:

$$\nabla E(\hat{\mathbf{D}}) = \hat{\mathbf{D}} \left[\sum_{i=1}^{N} \left(\frac{1}{b} \ln\left(\frac{S_i}{S_0}\right) + \mathbf{b}_i^T \hat{\mathbf{D}} \mathbf{b}_i \right) \cdot (\mathbf{b}_i \mathbf{b}_i^T) \right] \hat{\mathbf{D}}. \tag{26}$$

This allows us to design the appropriate gradient descent algorithm, with step length ϵ, in the Riemannian manifold $\mathscr{S}ym_6^+$:

$$\hat{\mathbf{D}}_{t+1} = \hat{\mathbf{D}}_t^{\frac{1}{2}} \exp\left(-\epsilon \cdot \hat{\mathbf{D}}_t^{\frac{1}{2}} \left[\sum_{i=1}^{N} \left(\frac{1}{b} \ln\left(\frac{S_i}{S_0}\right) + \mathbf{b}_i^T \hat{\mathbf{D}} \mathbf{b}_i\right) \cdot (\mathbf{b}_i \mathbf{b}_i^T)\right] \hat{\mathbf{D}}_t^{\frac{1}{2}}\right) \hat{\mathbf{D}}_t^{\frac{1}{2}}. \tag{27}$$

Minimizing the objective function $E(\hat{\mathbf{D}})$ in this way, it is possible to estimate $\hat{\mathbf{D}}$ in $\mathscr{S}ym_6^+$ from the diffusion signal. Since $\hat{\mathbf{D}}$ is isometrically equivalent to a 4th order tensor $\mathscr{D}^{(4)}$ with major and minor symmetries, $\mathscr{D}^{(4)}$ is guaranteed to have a positive diffusion profile. Finally we extract the totally symmetric part of $\mathscr{D}^{(4)}$ to compute the totally symmetric 4th order GDTI diffusion tensor $\mathscr{D}^{(4)s}$, which is then also guaranteed to have a positive diffusion profile.

3 A Ternary Quartic Approach for Symmetric Positive Semi-definite Fourth Order Diffusion Tensors

In this section, we revisit the problem of estimating a symmetric higher order Cartesian tensor with a positive diffusion profile from the GDTI model, using a polynomial approach. In this approach we consider the polynomial interpretation of HOTs instead of considering the algebra of HOTs, and look at a polynomial solution to the positivity problem. In particular, we consider 4th order GDTI diffusion tensors, where the diffusion function of such tensors can be seen as trivariate homogeneous polynomials of degree 4 in the coefficients of the gradient vector. Such polynomials are known as ternary quartics.

Polynomials form an alternate way of expressing the multi-linear form of HOTs. This expression was indicated in the original GDTI paper [4], but was used for applying the positivity constraint in [5]. To make the relationship between the coefficients of a HOT and the coefficients of a homogeneous polynomial more evident, the diffusion function of GDTI (Eq. (2)) was rewritten in [5] as:

$$D(\mathbf{g}) = \sum_{m+n+p=k} D_{m,n,p}\, g_1^m g_2^n g_3^p, \tag{28}$$

where $D_{m,n,p}$ are the coefficients of the kth order tensor $\mathscr{D}^{(k)}$ by a re-arrangement of the indices.

In this form, it is clear that the diffusion function, which was considered as the projection of the of a kth order HOT on to a unit sphere, is a trivariate homogeneous polynomial of degree k in the three coefficients of the unit gradient vector $\mathbf{g} = [g_1, g_2, g_3]^T$, where the coefficients of the polynomial are the coefficients of the HOT. Since $D(\mathbf{g})$ is a homogeneous polynomial of even degree, the problem of a positive diffusion profile on the unit sphere, $D(\mathbf{g}) > 0,\ \forall \mathbf{g} \in \mathbf{R}^3$ st. $||\mathbf{g}|| = 1$,

is equivalent to the problem of finding a polynomial $D(\mathbf{x}) > 0,\ \forall \mathbf{x} \in \mathbf{R}^3/\{\mathbf{0}\}$. This is exactly the same equivalence that was used in DTI, where the problem of positive diffusion from a second order tensor, $\mathbf{g}^T\mathbf{D}\mathbf{g} > 0,\ \forall \mathbf{g} \in S^2$, was recast as the problem of finding a positive definite second order tensor, $\mathbf{x}^T\mathbf{D}\mathbf{x} > 0,\ \forall \mathbf{x} \in \mathbf{R}^3/\{\mathbf{0}\}$, which entailed the Riemannian framework for $\mathscr{S}ym_3^+$. Therefore, in this section we consider a method of estimating the coefficients of a positive polynomial from the diffusion signal, to estimate a GDTI HOT with a positive diffusion profile.

3.1 Riemannian vs. Ternary Quartics: A Comparison

It is interesting to note at this juncture, when $k = 4$, how the Riemannian approach presented in the previous section compares to the polynomial formulation. When $k = 4$, the goal of the polynomial formulation, as we have just seen, is to find a trivariate homogeneous polynomial of degree 4, $D_4(\mathbf{x})$, where the coefficients of the polynomial are the coefficients of the 4th order GDTI diffusion tensor $\mathscr{D}^{(4)}$, such that:

$$D_4(\mathbf{x}) > 0, \quad \forall \mathbf{x} \in \mathbf{R}^3/\{\mathbf{0}\}. \tag{29}$$

In comparison, the Riemannian approach, using an isometric map, tries to find a symmetric 6D 2nd order tensor $\hat{\mathbf{D}}$ in $\mathscr{S}ym_6^+$:

$$\mathbf{c}^T\hat{\mathbf{D}}\mathbf{c} > 0, \quad \forall \mathbf{c} \in \mathbf{R}^6/\{\mathbf{0}\}, \tag{30}$$

where the coefficients of the totally symmetric 4th order GDTI diffusion tensor can be extracted from the coefficients of $\hat{\mathbf{D}}$. However, although, this quadratic form resembles the diffusion profile from a totally symmetric 4th order tensor, $\mathbf{b}^T\hat{\mathbf{D}}\mathbf{b}$ (Eq. (23)), estimating $\hat{\mathbf{D}}$ in $\mathscr{S}ym_6^+$ isn't equivalent to the problem of computing a 4th order GDTI diffusion tensor $\mathscr{D}^{(4)}$, with a positive diffusion profile. This can be seen from the isometrically equivalent inner product of the quadratic form:

$$\left\langle \mathbf{C}, \mathscr{D}^{(4)}\mathbf{C} \right\rangle > 0, \quad \forall \mathbf{C} \in \mathscr{S}ym_3/\{\mathbf{0}\}. \tag{31}$$

The positive diffusion profile constraint on the other hand only implies the condition:

$$\left\langle \mathbf{B}, \mathscr{D}^{(4)}\mathbf{B} \right\rangle > 0, \quad \text{where } \mathbf{B} = \mathbf{g} \otimes \mathbf{g}, \tag{32}$$

which can be seen in Eq. (22). Since the 2nd order tensor $\mathbf{B}$ in the diffusion profile is only of rank-1, it is rank deficient, whereas in general the 2nd order tensor $\mathbf{C}$, in the quadratic form would include both full rank, and rank deficient tensors. In other words, the positive quadratic form condition is much stronger than the

positive diffusion profile constraint. Therefore, although the positive quadratic form constraint would entail the positive diffusion profile constraint, the solutions found from this approach – the Riemannian approach, would only belong to a subset of all the solutions possible from only the positive diffusion profile constraint.

This can also be seen through examples, shown in [5, 6], by inspecting the isometric map in Eq. (16) which transforms a 4th order tensor into a 2nd order tensor. When this 6×6 matrix is positive definite it cannot represent valid totally symmetric 4th order tensors whose homogeneous polynomials are of the type $P(\mathbf{g}) = ag_1^4 + bg_2^4 + cg_3^4$, or $P(\mathbf{g}) = (ag_1^2 + bg_2^2)^2 + cg_3^4$, etc., because these require the matrix to be semi-definite [6]. Since, the Riemannian framework pushes such matrices away to an infinite distance from the estimation tensor $\hat{\mathbf{D}}$, the solutions found by the Riemannian estimation only form a subset of all possible solutions.

3.2 *Hilbert's Theorem on Non-negative Ternary Quartics*

We now return to the problem of estimating a non-negative trivariate homogeneous polynomial of degree k from the signal. A particular aspect of this problem has been addressed in [20], which describes a framework for estimating symmetric GDTI HOTs of any even order k and with a positive diffusion profile *on a unit sphere*. This paper proposes that any polynomial (the GDTI HOTs) that is non-negative on a unit sphere can be written as sums of squares of polynomials of lower order:

$$P^{(k)}(\mathbf{x}) = \sum_{i=1}^{M} Q_i^{(k/2)}(\mathbf{x}), \tag{33}$$

where k is even, $P^{(k)}(\mathbf{x})$ denotes a multi-variate polynomial of degree k, $\{Q_i^{(k/2)}(\mathbf{x})\}$ denote M multi-variate polynomials of degree $k/2$, and only an upper bound is known for M. Therefore, in [20], the authors propose to estimate the coefficients of the polynomials $Q_i^{(k/2)}(\mathbf{x})$ from the signal to estimate a polynomial $P^{(k)}(\mathbf{x})$ (or a GDTI HOT) with a non-negative diffusion profile.

Since M is not known exactly, the authors in [20] proceed by oversampling M, or rather densely sampling the space of possible polynomials of lower order $Q_i^{(k/2)}(\mathbf{x})$. It is claimed that increasing the density of the sampling increases the accuracy of the decomposition of $P^{(k)}(\mathbf{x})$. However, it also increases the number of unknown coefficients of the set $\{Q_i^{(k/2)}(\mathbf{x})\}$, which need to be estimated from the signal. The authors then propose heuristically measured approximations M' for M, for different values of k, from tests on synthetic data.

The problem of estimating non-negative trivariate polynomials when $k = 4$, or of estimating non-negative ternary quartics, presents a very interesting problem with a "complete" solution. In the case of ternary quartics, it can be shown that the entire space of non-negative polynomials over entire $\mathbf{R}^3$ (and not only over the

unit sphere), can be described by the sum of squares of quadratic polynomials. Examples in [21] of non-negative polynomials of degree $k > 4$ that cannot be written as sums of squares of lower order polynomials indicate that not all non-negative polynomials of arbitrary degree k can be decomposed into sums of squares of lower order polynomials. Hilbert's theorem, which identifies all the classes of non-negative multi-variate polynomials that can be always decomposed as sums of squares of lower order polynomials is also presented in [21].

In fact, Hilbert's theorem states that degree 4 trivariate polynomials that are non-negative and homogeneous, can always be written as a sum of squares of quadratic homogeneous polynomials, where the number of terms in the sum is also known and is exactly three ($M = 3$) [21]:

Theorem (Hilbert): If $P(x, y, z)$ is homogeneous, of degree 4, with real coefficients and $P(x, y, z) \geq 0$ at every $(x, y, z) \in \mathbf{R}^3$, then there are quadratic homogeneous polynomials f, g, h with real coefficients, such that:

$$P = f^2 + g^2 + h^2. \tag{34}$$

All other classes of non-negative polynomials that can be decomposed into sums of squares of lower order polynomials are all of degree less than four [21].

In this section, we, therefore, turn to Hilbert's theorem on non-negative, or positive semi-definite (PSD) ternary quartics, for a parameterization of the GDTI HOT when it is of order 4, to estimate diffusion HOTs with a non-negative diffusion profile. Since such tensors are symmetric and non-negative, these are known as symmetric positive semi-definite (SPSD) tensors. Based on Hilbert's theorem, [5] and [6] have proposed two different parameterizations of the 4th order tensor. A third parameterization was proposed in [22]. In this chapter, we review all three parameterizations, but follow through mainly with the method in [22].

As a final remark, we note that by adopting the polynomial formulation for the GDTI HOT, we have gained over the Riemannian framework proposed in the previous section from the fact that we address the exact problem of estimating a diffusion HOT with a positive diffusion profile, whereas the Riemannian approach addressed a more constrained problem. However, given the results on polynomials, namely Hilbert's theorem on ternary quartics, we concede to the Riemannian approach by the fact that we can only address the problem of a *non-negative* diffusion profile with the polynomial formulation, whereas the Riemannian approach addressed the *positive definite* diffusion profile constraint. However, we shall consider this a "negligible" loss, since in practice, due to numerical computations, we have never come across a diffusion profile that is exactly zero even along a single direction.

3.3 Estimating a SPSD Fourth Order Diffusion Tensor

The basic approach behind all three "ternary quartic" methods, [5, 6, 22], is the same. The idea is to consider the diffusion profile of a 4th order GDTI tensor as

a homogeneous trivariate polynomial in the coefficients of the gradient vector $\mathbf{g}$ (Eq. (28)), and to apply Hilbert's theorem on non-negative ternary quartics to rewrite it as a sum of squares of three quadratic homogeneous polynomials. Therefore, by estimating the coefficients of these quadratic homogeneous polynomials from the signal, it is possible to reconstruct the 4th order diffusion tensor by computing its coefficients from the coefficients of the quadratic forms, a process also known as the Gram-matrix approach [5, 23], such that the estimated 4th order tensor has a PSD diffusion profile. The three methods differ from each other in the way they parameterize the quadratic homogeneous polynomials to estimate their coefficients from the diffusion signal.

In [5], the diffusion profile of a 4th order GDTI tensor is written as:

$$D(\mathbf{g}) = (\mathbf{v}^T\mathbf{c}_1)^2 + (\mathbf{v}^T\mathbf{c}_2)^2 + (\mathbf{v}^T\mathbf{c}_3)^2, \tag{35}$$

$$= \mathbf{v}^T\mathbf{C}\mathbf{C}^T\mathbf{v}, \tag{36}$$

$$= \mathbf{v}^T\mathbf{G}\mathbf{v}, \tag{37}$$

where $\mathbf{v} = [g_1^2, g_2^2, g_3^2, g_1g_2, g_1g_3, g_2g_3]^T$ contains the monomials formed by the coefficients of the gradient vector $\mathbf{g}$, $\mathbf{v}^T\mathbf{c}_i$ are the three quadratic forms from Hilbert's theorem, and $\mathbf{G}$ is known as the Gram matrix. The column vectors $\mathbf{c}_i$ contain the coefficients of the quadratic forms, which have to be estimated from the signal, $\mathbf{C} = [\mathbf{c}_1|\mathbf{c}_2|\mathbf{c}_3]$ is a 6×3 matrix, which assembles these coefficients to compute the rank deficient or PSD 6×6 Gram matrix, which is used to compute the coefficients of the 4th order diffusion tensor from the coefficients of the quadratic forms.

The authors in [5] use the Eq. (36) to parameterize the ternary quartic decomposition by Hilbert's theorem, and estimate $\mathbf{C}$ from the DWIs, and compute the 4th order tensor from $\mathbf{G}$. However, this parameterization is problematic since it produces an infinite solution space, which can be seen by decomposing $\mathbf{C}$ into two blocks $\mathbf{C} = [\mathbf{A}, \mathbf{B}]^T$ where $\mathbf{A}$ and $\mathbf{B}$ are 3×3 matrices. Then $\mathbf{CO}$, for any 3×3 orthogonal matrix $\mathbf{O}$, also results in the same Gram matrix, since $\mathbf{CO}(\mathbf{CO})^T = \mathbf{C}\mathbf{C}^T = \mathbf{G}$. In other words, in this parameterization, $\mathbf{C}$ is unique only up to the equivalence class of orthogonal matrices $O(3)$.

In [5], the authors overcome this degenerate subspace issue by considering the $\mathbf{QR}$-decomposition (or $\mathbf{RQ}$-decomposition) of the 3×3 submatrix $\mathbf{A}$ of $\mathbf{C}$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ is an upper triangular matrix. This implies that $\mathbf{C} = [\mathbf{RQ}, \mathbf{B}]^T = [\mathbf{R}, \mathbf{B}\mathbf{Q}^T]^T\mathbf{Q}$. Therefore, $\mathbf{C}\mathbf{C}^T = [\mathbf{R}, \mathbf{B}\mathbf{Q}^T]^T\mathbf{Q} \cdot \mathbf{Q}^T[\mathbf{R}, \mathbf{B}\mathbf{Q}^T] = [\mathbf{R}, \mathbf{B}\mathbf{Q}^T]^T \cdot [\mathbf{R}, \mathbf{B}\mathbf{Q}^T]$, which effectively quotients out the orthogonal group from the computation of the Gram matrix $\mathbf{G}$.

In [6], the authors overcome this same issue in Eq. (36) by applying certain constraints on $\mathbf{C}$ from the properties of the Gram matrix, to remove the ambiguity of the class of orthogonal matrices $O(3)$. Since the rank of the Gram matrix is known a priori from Hilbert's theorem to be three, they identify and isolate the positive definite part of the PSD Gram matrix using a modified Iwasawa decomposition [24], which is then parameterized uniquely by a Cholesky decomposition. In other words,

they first collect the rank-3 positive definite part of $\mathbf{G}$ into a 3×3 matrix $\mathbf{W}$, and then decompose $\mathbf{W}$ using a Cholesky decomposition as $\mathbf{W} = \mathbf{L}\mathbf{L}^T$. This effectively equates the 3×3 matrix $\mathbf{A}$, from the paragraph above, where $\mathbf{C} = [\mathbf{A}, \mathbf{B}]^T$, to the triangular matrix with positive diagonal elements $\mathbf{L}$. In short, this procedure determines a unique $\mathbf{C}$ in the infinite space of solutions $\{\mathbf{CO}\}$ from the previous approach, and removes the ambiguity of the class of orthogonal matrices $O(3)$. Therefore, the authors in [6] effectively estimate $\mathbf{C} = [\mathbf{L}, \mathbf{B}]^T$ from the DWIs. Furthermore, the Cholesky decomposition also distinguishes $\mathbf{C}$ from $-\mathbf{C}$, although both result in the same Gram matrix. The authors then use this uniqueness property of $\mathbf{C}$ to design a spatial regularization of the field of estimated 4th order diffusion tensors.

Finally, we follow up in greater detail the third parameterization [22] using the ternary quartic decomposition. Essentially, using Eq. (35) to parameterize the Hilbert decomposition, we estimate the $\mathbf{c}_i$ directly from the DWIs and assemble these afterward to reconstruct $\mathbf{C}$. From there we follow the same procedure as the two other methods, and reconstruct the Gram matrix and compute the coefficients of the 4th order diffusion tensor.

From Hilbert's theorem on non-negative ternary quartics we write the diffusion function of a 4th order diffusion tensor as $D(\mathbf{g}) = \psi_1^2(\mathbf{g}) + \psi_2^2(\mathbf{g}) + \psi_3^2(\mathbf{g})$, where:

$$\psi_i(\mathbf{g}) = a_i g_1^2 + b_i g_2^2 + c_i g_3^2 + 2\alpha_i g_1 g_2 + 2\beta_i g_1 g_3 + 2\gamma_i g_2 g_3, \tag{38}$$

$$= [a_i, b_i, c_i, \sqrt{2}\alpha_i, \sqrt{2}\beta_i, \sqrt{2}\gamma_i] \tag{39}$$

$$\times [g_1^2, g_2^2, g_3^2, \sqrt{2}g_1 g_2, \sqrt{2}g_1 g_3, \sqrt{2}g_2 g_3]^T, \tag{40}$$

$$= \mathbf{x}_i^T \mathbf{v} \tag{41}$$

are the quadratic forms. Note that we have modified the form of the vector $\mathbf{v}$ by multiplying certain terms by $\sqrt{2}$, this is a minor difference in the notation convention from [5,6]. Each quadratic form is known if its six unknown coefficients in $\mathbf{x}_i$ can be estimated from the DWIs. Therefore, the diffusion profile can be written as a function of the unknowns to be estimated as:

$$D(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \mathbf{x}_1^T \mathbf{v}\mathbf{v}^T \mathbf{x}_1 + \mathbf{x}_2^T \mathbf{v}\mathbf{v}^T \mathbf{x}_2 + \mathbf{x}_3^T \mathbf{v}\mathbf{v}^T \mathbf{x}_3, \tag{42}$$

$$= [\mathbf{x}_1^T, \mathbf{x}_2^T, \mathbf{x}_3^T] \begin{bmatrix} \mathbf{v}\mathbf{v}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{v}\mathbf{v}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{v}\mathbf{v}^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} \tag{43}$$

$$= \mathbf{X}^T \mathbf{V} \mathbf{X}. \tag{44}$$

To estimate the unknown coefficients $\mathbf{x}_i$ of the homogeneous quadratic forms from a set of DWIs, we minimize the objective function based on the modified and linearized Stejskal-Tanner equation:

$$E(\mathbf{X}) = \frac{1}{2}\sum_{i=1}^{N}\left(\frac{1}{b}\log\left(\frac{S_i}{S_0}\right) + \mathbf{X}^T\mathbf{V}_i\mathbf{X}\right)^2, \tag{45}$$

where N is the number of DWIs and $\mathbf{V}_i$ corresponds to the monomials from the gradient direction $\mathbf{g}_i$. Although here we use the linearized form of the Stejskal-Tanner equation, it is equally possible to use the non-linear form. The gradient of the objective function with respect to the unknowns $\mathbf{X}$ is computed to be:

$$\nabla E(\mathbf{X}) = \sum_{i=1}^{N}\left(\frac{1}{b}\log\left(\frac{S_i}{S_0}\right) + \mathbf{X}^T\mathbf{V}_i\mathbf{X}\right)\left(\mathbf{V}_i + \mathbf{V}_i^T\right)\mathbf{X}. \tag{46}$$

We use the well known Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [25], a sophisticated quasi-Newton optimization algorithm for non-linear problems.

Finally we compute the 15 independent coefficients A_{ijkl} of the 4th order tensor $\mathscr{A}^{(4)}$ from the coefficients of the Gram matrix $\mathbf{G}$, by using Eq. (37), which equates $D(\mathbf{g})$, the multi-linear form of $\mathscr{A}^{(4)}$, to the quadratic form of the Gram matrix. We use a mapping very similar to the one presented in [5, 23], where the inverse mapping, i.e. $\mathbf{G}$ in terms of A_{ijkl} is given by:

$$\mathbf{G} = \begin{pmatrix} A_{xxxx} & a & b & \frac{1}{4}A_{xxxy} & \frac{1}{4}A_{xxxz} & d \\ a & A_{yyyy} & c & \frac{1}{4}A_{yyxy} & e & \frac{1}{4}A_{yyyz} \\ b & c & A_{zzzz} & f & \frac{1}{4}A_{zzxz} & \frac{1}{4}A_{zzyz} \\ \frac{1}{4}A_{xxxy} & \frac{1}{4}A_{yyxy} & f & \frac{1}{4}(A_{xyxy}-2a) & \frac{1}{8}(A_{xyxz}-4d) & \frac{1}{8}(A_{xyyz}-4e) \\ \frac{1}{4}A_{xxxz} & e & \frac{1}{4}A_{zzxz} & \frac{1}{8}(A_{xyxz}-4d) & \frac{1}{4}(A_{xzxz}-2b) & \frac{1}{8}(A_{xzyz}-4f) \\ d & \frac{1}{4}A_{yyyz} & \frac{1}{4}A_{zzyz} & \frac{1}{8}(A_{xyyz}-4e) & \frac{1}{8}(A_{xzyz}-4f) & \frac{1}{4}(A_{yzyz}-2c) \end{pmatrix}, \tag{47}$$

where $\{a, b, c, d, e, f\}$ are six free parameters that determine the rank of the matrix. In this case, since the rank of $\mathbf{G}$ is known to be three, the free parameters are determined from the construction of the Gram matrix, i.e. $\mathbf{G} = \mathbf{C}\mathbf{C}^T$. Therefore these can be used to compute the coefficients A_{ijkl}.

In comparison to the approach in [6], since we estimate all the coefficients of the three quadratic forms without any constraints, in effect we estimate 18 unknowns from which we recover the 15 unknowns of the 4th order diffusion tensor. This actually leaves us three degrees of freedom that can be applied as suitable constraints. Also this approach doesn't distinguish between $\mathbf{C}$ and $-\mathbf{C}$. However, since we only deal with the estimation problem of the 4th order diffusion tensor, this isn't important, since both $\mathbf{C}$ and $-\mathbf{C}$ give the same Gram matrix, and hence the same 4th order tensor. But if such were desired, the three degrees of freedom could be explored, to distinguish between $\mathbf{C}$ and $-\mathbf{C}$.

4 Experiments and Results

4.1 Synthetic Dataset

We conduct experiments on three datasets. First we consider a synthetic dataset based on a multi-tensor model (to represent multi-fiber crossings). For a single fiber profile we use the diagonal tensor $\mathbf{D} = \text{diag}(1{,}700, 300, 300) \times 10^{-6}\,\text{mm}^2/\text{s}$ and generate synthetic signals at a b-value of $3{,}000\,\text{s/mm}^2$. We estimate 4th order HOTs using the Riemannian and the "Ternary Quartic" (TQ) approaches and plot their ADCs. Further, since the maxima of the ADCs don't correspond to the fiber directions, we also compute the diffusion ensemble average propagators (EAPs) $P(\mathbf{r}) = \int (S_i(\mathbf{q})/S_0) \exp\left(-2\pi i \mathbf{q}^T \mathbf{r}\right) d\mathbf{q}$, from the estimated 4th order tensors [26, 27].

We visually compare the Riemannian approach, which guarantees a positive definite diffusion profile but solves a more constrained problem, to the Ternary Quartic approach, which guarantees only a positive semi-definite diffusion profile but solves the problem in the correct space. The diffusion profiles of the estimated 4th order GDTI tensors and the EAPs computed thereof are presented in Fig. 1. We notice that the ADCs and the EAPs of the Ternary Quartic approach are somewhat sharper than the Riemannian counterparts. We surmise that this is due to the fact that the Riemannian approach cannot estimate certain types of 4th order tensors that can have non-negative diffusion profiles, since these tensors require to have a semi-definite representation in the symmetric 6D 2nd order tensor formulation used by the Riemannian estimation. Such semi-definite 6D 2nd order tensors are, however, pushed to an infinite distance from the estimation tensor by the Riemannian metric. Nonetheless, the overall angular structure of the two methods remain comparable.

4.2 Biological Phantom Dataset

Next we conduct an experiment on a biological phantom data that was produced from excised rat spinal cords. Only two cords were used to create a fiber crossing configuration with known physical directions. The biological phantom [28] was created at the McConnell Brain Imaging Center (BIC), McGill University, Montréal, Canada. MR images were acquired on a 1.5T Sonata MR scanner using a knee coil. It was created from two excised Sprague-Dawley rat spinal cords embedded in 2 % agar. The acquisition was done with a single-shot spin-echo planar sequence with twice-refocused balanced gradients, designed to reduce eddy current effects. The dataset was acquired with 90 gradient directions, on a single q-shell with a b-value of $3{,}000\,\text{s/mm}^2$, $q = 0.35\ \mu\text{m}^{-1}$, TR = 6.4 s, TE = 110 ms, FOV $360 \times 360\,\text{mm}^2$, 128×128 matrix, 2.8 mm isotropic voxels and four signal averages per direction. The SNR of the S_0 image was estimated to be approximately 70 for the averaged phantom, and around 10 for the cord at b-value of $3{,}000\,\text{s/mm}^2$.

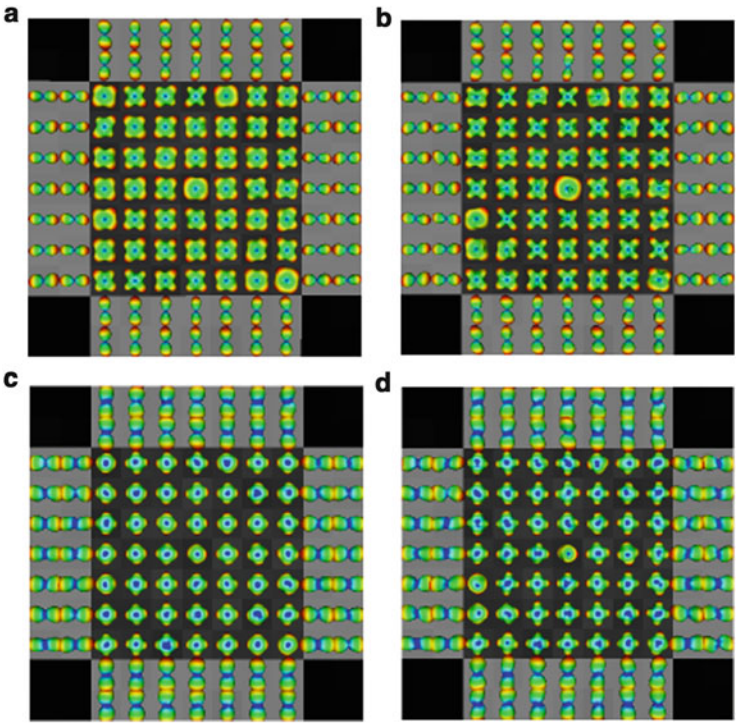


Fig. 1 Synthetic dataset. Comparing the diffusion profiles and the EAPs from the Riemannian approach and the Ternary Quartic approach. (**a**) ADC Riemannian. (**b**) ADC Ternary Quartic. (**c**) EAP Riemannian. (**d**) EAP Ternary Quartic. The Riemannian approach guarantees positive diffusion, but solves a more constrained problem. The Ternary Quartic approach guarantees only a positive semi-definite diffusion, but solves the problem in the correct space

In this experiment we estimate 4th order GDTI diffusion tensors from the phantom dataset using both the Riemannian approach and the TQ approach. We then compute the EAPs from the tensors using the methods in [26,27] to validate the coherence of their geometry with the known layout of the phantom and to see if it is possible to infer the underlying fiber bundle directions. For the sake of comparison we also present the result of the orientation distribution function (ODF) computed from the analytical q-ball estimation technique in [29], which is an angular marginal distribution of the true and unknown EAP under a mono-exponential decay model that corresponds to the GDTI model. The ODFs were directly estimated from the signal.

The results are presented in Fig. 2. The geometry of the EAPs computed from the 4th order tensors estimated using both the methods are coherent with the underlying phantom model, and also agree with the geometry of the ODFs. It is interesting to note that since the ODFs are angular marginal distributions of the true EAPs, the radial information of the true EAPs has been marginalized out by a radial

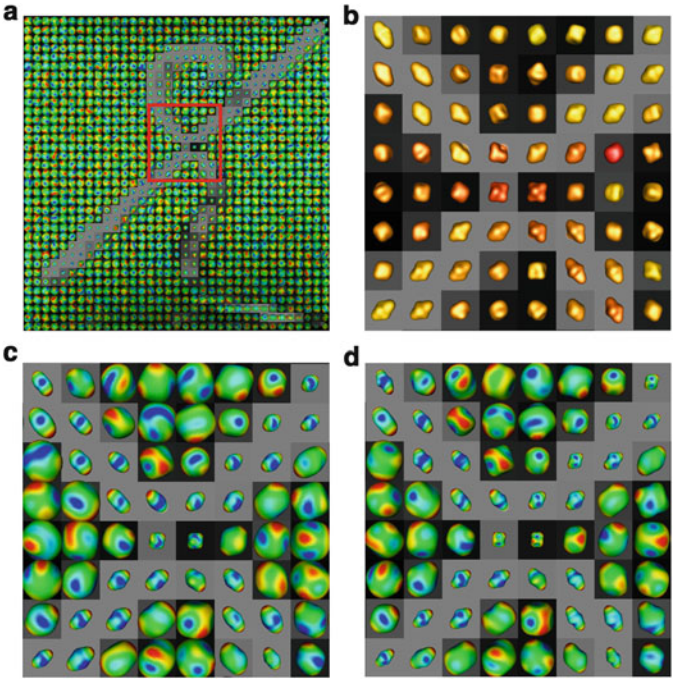


Fig. 2 Biological phantom dataset. (**a**) The layout of the phantom created using two excised rat spinal cords. (**b**) ODFs estimated from the signal as reference geometry. (**c**) EAPs computed from 4th order tensors estimated using the Riemannian approach. (**d**) EAPs computed from 4th order tensors estimated using the Ternary Quartic approach. The EAPs were evaluated at the constant probability radius of $|\mathbf{r}| = 17\ \mu\text{m}$

integration. Therefore, although the ODFs' angular structures resemble the angular structures of the EAPs computed from the 4th order tensors, the ODFs do not reveal anything about the magnitude of diffusion due to the heterogeneous structure of the underlying tissue. This is visible in the EAPs computed from the tensors from the size or volume of the displacement probability at a constant displacement radius. Also, by comparing (c) and (d) in Fig. 2, again the EAPs from the TQ method look sharper than the Riemannian counterparts.

4.3 In Vivo Human Dataset

Finally we conduct experiments on an in vivo human cerebral dataset. This dataset was acquired on a 1.5T scanner using 41 gradient directions, with a b-value of $700\ \text{s/mm}^2$ with $\text{TR} = 1.9\ \text{s}$, $\text{TE} = 93.2\ \text{ms}$, 128×128 image matrix, 60 slices, with voxel dimensions of $1.875 \times 1.875 \times 2\ \text{mm}$. This dataset is from a public HARDI database that can be found in [30].

Table 1 Real dataset. The estimated diffusion functions from 249,352, 4th order GDTI tensors checked for positive diffusion profile on a set of 81 pairs of directions distributed evenly on a sphere. The Ternary Quartic and the Riemannian approaches are the only methods, which guarantee positive diffusion

(81 dirs)	LS	SH	RM	TQ
Positive	181,757	249,263	249,352	249,352
Negative	67,595	89	0	0

We consider two regions of interest (ROIs) with 249,352 and 987 voxels respectively. For the 249,352 voxels we compute the diffusion profiles of the tensors and test for positive/non-negative diffusion along 81 directions distributed evenly on a hemisphere. For the 987 voxels we compare the estimation time of the methods, since the positivity constraint implies increased computational complexity.

For the positive/non-negative diffusion experiment, we test four approaches. First we consider the standard Euclidean least squares approach (LS). Then we also test a method based on spherical harmonics (SH). Since SHs of the same rank are bijective to Cartesian tensors of the same order, we first estimate real and symmetric SHs of rank 4 from the signal and then transform them to the tensor basis to obtain 4th order HOTs. And finally we consider the two proposed methods of this chapter, namely the Riemannian approach and the TQ approach. The LS approach and the SH to HOT approach don't consider any constraints, although the SH approach includes Laplace-Beltrami regularization [31] to account for some signal noise.

The results of this experiment are displayed in Table 1. The Riemannian (RM) approach and the TQ approach are the only two that estimate 4th order diffusion tensors with positive diffusion profiles. The LS approach, as known, estimates tensors with lots of negative diffusion directions. Although the SH to HOT method includes regularization, clearly that is insufficient to guarantee positive diffusivity. Positive diffusivity is only achieved when it is applied explicitly by either the Riemannian approach or the TQ approach. In this experiment, we also tested for zero diffusion and found that both the Riemannian method and the TQ method always estimated tensors with strict positive diffusion profiles. Although the TQ method only applies a non-negative constraint, clearly due to numerical computations it is highly improbable to estimate tensors with exactly zero diffusion.

Although, the positivity constraint, applied using either the Riemannian approach or the TQ approach, clearly performs well, it also implies an added computational load. To get an idea of the additional computational complexity, we compare the estimation time of the two – Riemannian & TQ – approaches with the standard and linear LS approach on an ROI of the in vivo dataset with 987 voxels. The estimation times are displayed in Table 2. The computations were conducted on a Dell D630 Latitude laptop with Intel(R) Core(TM)2 Duo CPU @ 2.20 GHz and 2 GB RAM. The linearity and efficiency of the LS method is in fact one of its main supporting factors. However, the increased estimation time due to the complexity of the positivity constraint is still tractable.

Table 2 Real dataset. Comparison of the time for estimating 987, 4th order diffusion tensors that are visualized in Fig. 3.

(987 tensors)	LS	Rm	TQ
Estimation	6s	35s	102s

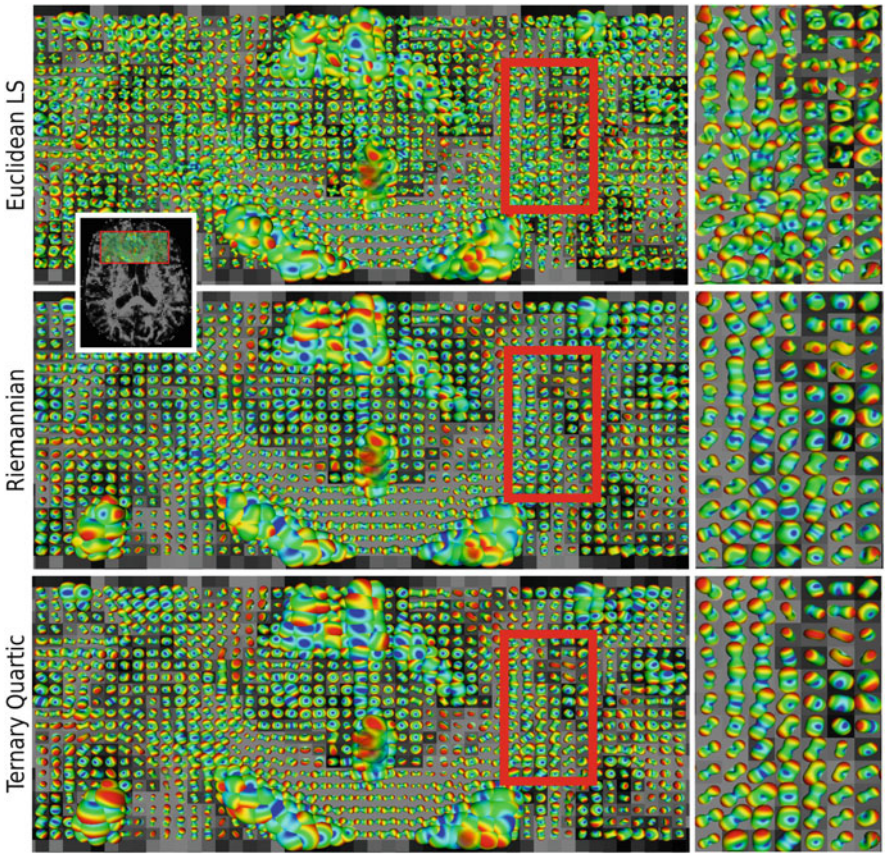


Fig. 3 In-vivo human cerebral dataset. Effects of the non-negative and the positive definite constraints that are guaranteed by the Ternary Quartic approach and the Riemannian approach are evaluated on the EAPs computed from the estimated tensors. EAPs computed from tensors estimated using the Euclidean LS approach, which doesn't consider any constraints, are shown for comparison. No spatial regularization was used. The improvement in the results is only due to the non-negativity constraints

Finally, we conclude the experiments, by computing the EAPs from tensors estimated using both the Riemannian method and the TQ method from the in vivo human dataset (using [26,27]). For comparison we include the EAPs computed from tensors estimated using the LS method (using [26, 27]). The results are presented in Fig. 3, where a region of interest on an axial slice is shown. What stands out

prominently from Fig. 3 is the increased spatial regularity in the results of the Riemannian and the TQ methods when compared to the LS method. However, no spatial regularization was used. Only the positivity constraint was employed, using the two methods, for estimating the 4th order tensors. Clearly, the positivity constraint renders the estimation of the tensors much more robust to signal noise and improves the results. This indicates the importance of the positivity constraint.

5 Discussion and Conclusion

In this chapter we considered the problem of estimating 4th order diffusion tensors with a positivity constraint from the GDTI model. In GDTI Cartesian tensors of order higher than two were used to attain greater accuracy in the modeling of complex shaped ADCs. GDTI HOTs of order k were assumed to be symmetric since only their projections along vectors were used in the ADC modeling, and were assumed to be of even order since negative diffusion is non-physical. However, in spite of this design, the standard method for estimating GDTI HOTs from the signal, namely the least squares approach doesn't guarantee an estimated HOT with a positive diffusion profile. Least squares estimation, although linear and efficient can result in HOTs with negative diffusion profiles.

We reviewed two different approaches for estimating 4th order GDTI diffusion tensors with positive diffusion profiles and non-negative diffusion profiles respectively. In the first method, we considered the algebra of 4th order tensors to map symmetric 3D 4th order tensors to symmetric 6D 2nd order tensors. We then applied the Riemannian framework for the space of $\mathscr{S}ym_6^+$, to estimate 4th order diffusion tensors with strictly positive or positive definite diffusion profiles. In the second method, we considered the polynomial interpretation of the multi-linear form of HOTs, to reformulate the problem of estimating a HOT as a problem of estimating a polynomial. In the case of 4th order diffusion tensors, we were able to use Hilbert's theorem on non-negative ternary quartics to parameterize 4th order tensors as a sum of squares of quadratic forms. By estimating the coefficients of the quadratic forms, we were able to reconstruct 4th order diffusion tensors with non-negative diffusion profiles from the signal.

The Riemannian method we proposed, ensures a positive definite diffusion profile, but solves a problem more constrained than implied by the model. This can be understood from the fact that the 3D 4th order tensors were estimated in $\mathscr{S}ym_6^+$, as 6D 2nd order tensors, which implies that the Riemannian method ensures that the multi-linear form of the 4th order tensor is positive definite for all symmetric 3D 2nd order tensor. However, the GDTI model requires that the multi-linear form of the 4th order tensor needs to be positive definite for only 3D 2nd order tensors of maximal rank one. Therefore, the Riemannian method ensures a positive diffusion profile, but the solution space is more constrained than the true solution space.

The second method we proposed – the Ternary Quartic method solves the problem in the correct space due to the appropriate polynomial parameterization.

However, since the known polynomial results, i.e. Hilbert's theorem on ternary quartics, only guarantee non-negativity, this method considers a theoretically weaker problem of a positive semi-definite diffusion profile. But this method also uses the Euclidean metric, which, as has been suggested in [7], is perhaps better suited for computing with diffusion data than the affine invariant Riemannian metric.

From the implementation and the results, we found that the shape of the ADCs and EAPs computed from tensors estimated using the Riemannian method to be similar to the shape of the ADCs and EAPs computed from tensors estimated using the Ternary Quartic method. We did, however, remark a swelling in the shapes of the tensors estimated using Riemannian method, which we suspect was the result of the over constraint. A more detailed analysis is, therefore, necessary to identify the sub-space spanned by the Riemannian approach, and also to quantify the impact of this sub-space on the estimated results. Finally, in the tests for negative diffusion profiles, we never came across zero diffusion from tensors estimated using the Ternary Quartic method, which is probably due to numerical computations. Therefore, we concluded that the concession of the weaker non-negativity constraint to be negligible in practice.

We conducted tests on a biological phantom with a known layout to evaluate whether it was possible to infer the underlying fiber directions from the geometry of the EAPs computed from the tensors estimated using the two approaches. Our experiments indicated that this could be answered in the affirmative and that the geometry of the EAPs computed from the tensors estimated using the Riemannian framework and the Ternary Quartic approach could reveal the underlying fiber directions. We also experimented on in-vivo human cerebral data using both the Riemannian framework and the Ternary Quartic approach to motivate the need for a positive or non-negative diffusion profile constraint. The experiments clearly indicated the gains of applying such constraints. Finally, we also presented the computation time to evaluate the increased complexity, and found this to be tractable.

Acknowledgements The authors are partially supported by the NucleiPark research project (ANR Program "Maladies Neurologique et maladies Psychiatriques") and the France Parkinson Association.

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson. B **103**, 247–254 (1994)
2. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**, 259–267 (1994)
3. Stejskal, E.O.: Use of spin echoes in a pulsed magnetic-field gradient to study anisotropic, restricted diffusion and flow. J. Chem. Phys. **43**, 3597–3603 (1965)
4. Ozarslan, E., Mareci, T.H.: Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. Magn. Reson. Med. **50**, 955–965 (2003)

5. Barmpoutis, A., Jian, B., Vemuri, B.C.: Symmetric positive 4th order tensors & their estimation from diffusion weighted MRI. In: Information Processing in Medical Imaging (IPMI 2007), Kerkrade, pp. 308–319 (2007)
6. Barmpoutis, A., Hwang, M.S., Howland, D., Forder, J.R., Vemuri, B.C.: Regularized positive-definite fourth-order tensor field estimation from DW-MRI. NeuroImage **45**, 153–162 (2009)
7. Pasternak, O., Sochen, N., Basser, P.J.: The effect of metric selection on the analysis of diffusion tensor MRI data. NeuroImage **49**, 2190–2204 (2010)
8. Qi, L., Yu, G., Wu, E.X.: Higher order positive semidefinite diffusion tensor imaging. SIAM J. Imaging Sci. **3**, 416–433 (2010)
9. Milne, T., Ghosh, A., Deriche, R.: Constrained diffusion kurtosis imaging using ternary quartics and MLE. In: Computational Diffusion MRI Workshop (CDMRI), 2012, MICCAI, Nice (2012)
10. Fletcher, P., Joshi, S.: Principal geodesic analysis on symmetric spaces: statistics of diffusion tensors. In: Proceedings of the Computer Vision Approaches to Medical Image Analysis, Prague (2004)
11. Moakher, M.: A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl. **26**, 735–747 (2005)
12. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. Int. J. Comput. Vis. **66**, 41–66 (2006). A preliminary version appeared as INRIA Research Report 5255, July 2004.
13. Lenglet, C.: Geometric and variational methods for diffusion tensor MRI processing. PhD thesis, Universite de Nice – Sophia Antipolis (2006)
14. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magn. Reson. Med. **56**, 411–421 (2006). PMID: 16788917.
15. Moakher, M.: Fourth-order cartesian tensors: old and new facts, notions and applications. Q. J. Mech. Appl. Math. **61**, 181–203 (2008)
16. Moakher, M.: The algebra of fourth-order tensors with application to diffusion MRI. In: Laidlaw, D., Weickert, J. (eds.) Visualization and Processing of Tensor Fields. Mathematics and Visualization, pp. 57–80. Springer, Berlin/Heidelberg (2009). 10.1007/978-3-540-88378-4_4
17. Itskov, M.: On the theory of fourth-order tensors and their applications in computational mechanics. Comput. Methods Appl. Mech. Eng. **189**, 419–438 (2000)
18. Ghosh, A., Descoteaux, M., Deriche, R.: Riemannian framework for estimating symmetric positive definite 4th order diffusion tensors. In: Medical Image Computing And Computer-Assisted Intervention – MICCAI. Volume 5241 of Lecture Notes in Computer Science, New York, pp. 858–865. Springer (2008)
19. Basser, P.J., Pajevic, S.: Spectral decomposition of a 4th-order covariance tensor: applications to diffusion tensor MRI. Signal Process. **87**, 220–236 (2007)
20. Barmpoutis, A., Vemuri, B.C.: A unified framework for estimating diffusion tensors of any order with symmetric positive-definite constraints. In: Proceedings of ISBI10: IEEE International Symposium on Biomedical Imaging, Rotterdam, pp. 1385–1388 (2010)
21. Rudin, W.: Sums of squares of polynomials. Am. Math. Mon. **107**, 813–821 (2000)
22. Ghosh, A., Moakher, M., Rachid, D.: Ternay quartic approach for positive 4th order diffusion tensors revisited. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, pp. 618–621 (2009)
23. Powers, V., Reznick, B.: Notes towards a constructive proof of Hilbert's theorem on ternary quartics. In: Quadratic Forms and Their Applications (Dublin 1999). Contemporary Mathematics, vol. 272. American Mathematical Society, Providence (2000)
24. Jian, B., Vemuri, B.C.: Metric learning using Iwasawa decomposition. In: IEEE International Conference on Computer Vision, Rio de Janeiro **0**, pp. 1–6 (2007)
25. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes in C: The Art of Scientific Computing, 2nd edn. Cambridge University Press, Cambridge/New York (1992)
26. Özarslan, E., Vemuri, B.C., Mareci, T.H.: Fiber orientation mapping using generalized diffusion tensor imaging. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Arlington, pp. 1036–1039 (2004)

27. Ghosh, A., Deriche, R.: Fast and closed-form ensemble-average-propagator approximation from the 4th-order diffusion tensor. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, Rotterdam, pp. 1105–1108 (2010)
28. Campbell, J., Siddiqi, K., Rymar, V., Sadikot, A., Pike, B.: Flow-based fiber tracking with diffusion tensor q-ball data: validation and comparison to principal diffusion direction techniques. NeuroImage **27**, 725–736 (2005)
29. Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N.: Reconstruction of the orientation distribution function in single and multiple shell q-ball imaging within constant solid angle. Magn. Reson. Med. **64**, 554–566 (2010)
30. Poupon, C., Poupon, F., Allirol, L., Mangin, J.F.: A database dedicated to anatomo-functional study of human brain connectivity. In: Twelfth Annual Meeting of the Organization for Human Brain Mapping (HBM), Florence (2006)
31. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Apparent diffusion coefficients from high angular resolution diffusion imaging: estimation and applications. Magn. Reson. Med. **56**, 395–410 (2006)

Riemann-Finsler Geometry for Diffusion Weighted Magnetic Resonance Imaging

Luc Florack and Andrea Fuster

Abstract We consider Riemann-Finsler geometry as a potentially powerful mathematical framework in the context of diffusion weighted magnetic resonance imaging. We explain its basic features in heuristic terms, but also provide mathematical details that are essential for practical applications, such as tractography and voxel-based classification. We stipulate a connection between the (dual) Finsler function and signal attenuation observed in the MRI scanner, which directly generalizes Stejskal-Tanner's solution of the Bloch-Torrey equations and the diffusion tensor imaging (DTI) model inspired by this. The proposed model can therefore be regarded as an extension of DTI. Technically, reconstruction of the (dual) Finsler function from diffusion weighted measurements is a fairly straightforward generalization of the DTI case. The extension of the Riemann differential geometric paradigm for DTI analysis is, however, nontrivial.

1 Introduction

Diffusion weighted magnetic resonance imaging (dwMRI) has become a standard MRI technique for in vivo imaging of apparent water diffusion processes in fibrous tissue (for an introductory tutorial, cf. Hagmann et al. [1]). Clinical use of dwMRI is hampered by the fact that radically new approaches and abstract representations are required for its analysis. Examples are rank-2 symmetric positive-definite tensor representations in diffusion tensor imaging (DTI), pioneered by Basser, Mattiello and Le Bihan et al. [2–8] and explored by many others [9–20], higher order fully symmetric tensor representations [21–27] and spherical harmonic representations in

L. Florack (✉) • A. Fuster
Eindhoven University of Technology, PO Box 513, NL-5600 MB Eindhoven, The Netherlands
e-mail: L.M.J.Florack@tue.nl; A.Fuster@tue.nl

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_8,

high angular resolution diffusion imaging (HARDI) [28–32], and SE(3) Lie group representations [33–35].

In this chapter we concentrate on an extension of the Riemannian paradigm [16, 36], used in the context of DTI, in order to account explicitly for the unconstrained number of local directional degrees of freedom of general dwMRI representations. *Riemann-Finsler geometry* appears to be ideally suited for this purpose, as has already been hinted upon in earlier work [22, 37–41]. However, foregoing work is either driven by heuristics or merely scratches the surface of Riemann-Finsler geometry. For instance, no rigorous connection between the pivotal Finsler function and the physics of dwMRI acquisition has yet been proposed.

More specifically, Melonakos et al. [41] have pioneered Finsler geometry in the context of contours, only briefly touching upon application in dwMRI. Astola et al. [37–40] have applied Finsler geometry, and in particular geodesic tractography, to dwMRI using a fully symmetric fourth order tensor model. Florack et al. [22] have proposed a tensor representation of arbitrary order, discussing operational issues such as spatial and angular regularization. The Cartan geometric approach developed by Duits et al. [33–35] likewise appears intimately related to the theory outlined in this chapter. In all of the above cases the exact connection between Riemann-Finsler geometry and dwMRI is deemphasized, while applications are typically limited to contour detection or tractography.

Our primary goal is to provide a generic model for dwMRI, with potential applications beyond tractography, which manifestly incorporates the Riemannian paradigm for DTI as a limiting case. Secondly, we wish to convey the gist of Riemann-Finsler geometry without dodging mathematical details that are necessary for algorithmic implementation. This does not imply that our treatment of the subject will be self-contained; for a thorough understanding one will find it necessary to consult additional sources. The books by Bao et al. [42] and Rund [43] are especially recommended. Shen and Mo provide additional insight [44, 45]. We hope that our overview will encourage researchers to further contribute to a systematic study and practical application of Riemann-Finsler geometry in the context of dwMRI (and elsewhere).

Riemann-Finsler geometry has its roots in Riemann's "Habilitation" [46]. Riemann focused on a special case, nowadays known as Riemannian geometry. Important (pseudo-Riemannian) application areas, such as Maxwell's theory and Einstein's theory of general relativity, greatly contributed to its popularity. The general case was taken up by Finsler in his PhD thesis [47], and subsequently by Cartan [48] (who was the first to refer to it as "Finsler geometry"), and others.

Although potentially much more powerful, Riemann-Finsler geometry has not yet become nearly as popular as its Riemannian counterpart. To some extent this may be explained by its rather mind-boggling technicalities and heavy computational demands. This should no longer withhold practitioners in our technological era, for both symbolic as well as large-scale numerical manipulations can be readily performed on state-of-the-art computers. Progress in enabling technologies, such as compressed sensing for fast imaging [49], are also likely to contribute to practical feasibility of dwMRI, yet we believe that the major hurdle is still methodological.

2 Theory

2.1 Diffusion Weighted MRI

Recall the Stejskal-Tanner signal attenuation formula in the spin-echo experiment on spin diffusion in an isotropic medium [50] (cf. also [51–55]):

$$\ln \frac{A(2\tau)}{A(0)} = -\gamma^2 D\delta^2(\Delta - \frac{1}{3}\delta)G^2\,, \tag{1}$$

in which γ is the gyromagnetic ratio of hydrogen, δ the duration of a diffusion-sensitizing gradient pulse (with $\delta < \tau$), Δ the time between a pair of balanced gradient pulses, and G the gradient magnitude. The echo occurs at time 2τ after the onset of the first gradient pulse, and the formula represents the relative signal loss due to diffusion of water molecules over a time interval τ between the 90° pulse and the 180° pulse, a process characterized by the diffusion coefficient D. The positive factor by which D is multiplied on the right hand side of the above expression is known in the trade as the "b-factor" (an allusion to Le Bihan): $b \equiv \gamma^2\delta^2(\Delta - \frac{1}{3}\delta)G^2$.

Brain white matter consists mostly of water (>70 %), but diffusion turns out to be anisotropic as a result of its fibrous architecture, facilitating diffusion along axonal fibers relative to transverse directions. The Stejskal-Tanner experiment inspired Moseley, Basser, Le Bihan, and others, [2–8] to capture this anisotropy in terms of a symmetric positive definite rank-2 diffusion tensor (with components D^{ij}, $i, j = 1, 2, 3$, relative to a coordinate basis) as opposed to the scalar D. This technique is the basis of *diffusion tensor imaging* (DTI).

In order to connect to the mathematical notation in the remainder of this chapter we will denote the signal as a function of position x and the applied normalized diffusion-sensitizing gradient $q = \gamma\delta G$:

$$S(x, q, \tau^*) = S_0(x) \exp\left(-\tau^* H^2(x, q)\right)\,. \tag{2}$$

Here τ^* denotes a time constant related to Δ and δ (in Stejskal-Tanner's scheme we have $\tau^* = \Delta - \delta/3$, cf. Sinnaeve [55] for alternative schemes and associated time constants). Furthermore, the so-called Hamiltonian[1] $H(x, q)$ generalizes the quadratic form encountered in the DTI case, in which it assumes the form[2]

$$H^2_{\mathrm{DTI}}(x, q) = D^{ij}(x) q_i q_j\,. \tag{3}$$

[1] We neglect, but do not a priori exclude, a mild dependence of $H(x, q)$ on τ^*.

[2] We use Einstein's summation convention throughout, i.e. explicit summation symbols, such as $\sum_{i,j=1}^{3}$ on the r.h.s. of Eq. (3), are suppressed for pairs of identical upper and lower indices.

In general, $H(x,q) \neq H_{\mathrm{DTI}}(x,q)$, but a strong analogy with DTI remains in the form of a homogeneity condition[3] in q-space, viz. we shall require that

$$H^2(x,q) = D^{ij}(x,q)q_i q_j \,, \tag{4}$$

in which the coefficients are zero-homogeneous, i.e.

$$D^{ij}(x,\lambda q) = D^{ij}(x,q) \tag{5}$$

for all $\lambda \neq 0$. This expresses our hypothesis that $\ln(S(x,q,\tau^*)/S_0(x))$ scales quadratically in the *magnitude* of the diffusion-sensitizing gradient, but, unlike DTI, is not necessarily a quadratic form. This assumption is approximately correct for certain ranges of (q,τ^*), and encompasses the validity domain of DTI.

By virtue of homogeneity and mirror symmetry one may, in Eq. (5), think of q as a point on the projective plane, or on the unit sphere with antipodal points identified. Homogeneity also implies that the "highly anisotropic" diffusion tensor $D^{ij}(x,q)$ does *not* in fact—within the domain of validity of our extended model—depend on acquisition details, such as the magnitude of the applied gradients. That is, it is intended to capture tissue intrinsic properties (probed along a certain direction), just like the classical, "mildy anisotropic" diffusion tensor $D^{ij}(x)$. (This does *not* hold for $H(x,q)$ and some related functions that will be introduced below, which do depend on the magnitude of q, and thus on the entire experimental setup.)

Note that the number of degrees of freedom contained in the DTI tensor coefficients $D^{ij}(x)$ at any given point x equals 6 (the number of independent components of a symmetric 2-tensor in 3 dimensions), whereas there are, a priori, ∞ degrees of freedom in $D^{ij}(x,q)$, one for each position in space and each point on the projective plane. Also note that Eq. (2) relies on a mono-exponential signal decay; in this sense it "naturally" extends DTI. It complements alternative DTI refinements, such as multi-compartment models [56]. Our homogeneity condition,

$$H(x,\lambda q) = |\lambda| H(x,q) \qquad \text{for all } \lambda \in \mathbb{R}, \tag{6}$$

also distinguishes our model from diffusional kurtosis imaging (DKI), cf. [57]. (A comparison of our model with these models as well as other, HARDI-like schemes, in relation to their respective validity domains, remains to be made.)

2.2 The Riemannian Paradigm

The Riemannian paradigm was introduced by O'Donnell et al. [36] and by Lenglet et al. [16] in the context of DTI. In its original formulation it identifies the diffusion

[3] A function $f(z)$ is called homogeneous of degree r if it satisfies $f(\lambda z) = \lambda^r f(z)$ for all $\lambda > 0$. According to Euler's theorem such a function obeys the first order partial differential equation $z^i \partial f(z)/\partial z^i = r f(z)$.

tensor $D^{ij}(x)$, recall Eqs. (2) and (3), up to a constant proportionality factor, with the dual (or inverse) Riemann metric tensor $g^{ij}(x)$:

$$\tau^* D^{ij}(x) = g^{ij}(x)\,. \tag{7}$$

This defines a Riemannian manifold in which a relatively increased directional diffusion observed along some curve is tantamount to a shortening of Riemannian path length. In this way the problem of tractography can be restated as the problem of finding certain[4] shortest paths (via geodesic equations), or related to level set methods for distance functions induced by geodesic congruences (via Hamilton-Jacobi equations). The motivation for Eq. (7) is heuristic, cf. [58, 59] for conformal adaptations of the metric, arguing for a nontrivial local scaling factor.

Due to its limited angular resolution DTI can only handle mild anisotropies that are believed to be induced by "single fiber coherence", i.e. a local alignment of axonal fibers forcing anisotropy to be more or less axially symmetric, with one dominant diffusion eigenaxis along the fibers (and two minor eigenaxes perpendicular to the fibers). Due to complex fiber architecture in significant parts of the brain, such as fiber crossings, observed diffusivity patterns are highly anisotropic, rendering the DTI hypothesis invalid in such cases. On the positive side, the same limitation (viz. of a priori limited angular resolution) contributes to robustness, especially if a redundant set of diffusion weighted measurements is used for DTI reconstruction.

If we drop the quadratic restriction, Eq. (3), we can invoke the powerful machinery of Riemann-Finsler geometry in a way that mimics the Riemannian paradigm for DTI, viz. recall Eq. (4) and identify the coefficients in this equation with the so-called *dual Riemann-Finsler metric tensor*:

$$\tau^* D^{ij}(x,q) = g^{ij}(x,q)\,. \tag{8}$$

Clearly this is at best an approximation of reality due to mono-exponential decay, Eq. (2), and the homogeneity hypothesis on the physical scaling behavior of the Hamiltonian, Eq. (6). The conditions under which this approximation is realistic are deemphasized here, but will need to be made explicit in future work (cf. Basser [6] for a discussion in the DTI case, to some extent applicable to the general case as well). Suffice it to say that, by construction, the domain of validity certainly reaches beyond that of DTI, which arises in the limiting scenario of mild anisotropy $D^{ij}(x,q) \to D^{ij}(x)$.

In the rest of this chapter we consider the basics of Riemann-Finsler geometry and point out its theoretical relevance for tractography and voxel classification.

[4]Please note that the Riemannian paradigm does *not* stipulate that geodesics are biologically meaningful tracts, cf. Astola et al. [10] for a connectivity criterion that could be used for a deterministic or probabilistic labelling of biologically plausible neural tracts among all possible geodesic tracts. Indeed, in a geodesically complete space *any* two points can be connected by at least one geodesic.

The Hamiltonian framework appears to be most directly related to the physics of dwMRI. In DTI this is reflected by the fact that it is the *inverse* of the diffusion tensor that defines the Riemann metric tensor. The Riemann metric tensor itself is equivalently captured by a (limiting case of a) so-called *Finsler function*, which, in its most general form, constitutes the pivot of Riemann-Finsler geometry. Let us therefore start with the axiomatics of the Finsler function.

2.3 The Finsler Function

Recall that the geometric paradigm for DTI hinges on Riemannian geometry, Eq. (7), stipulating that the diffusion tensor is proportional to the dual Riemann metric tensor $g^{ij}(x)$, with 6 degrees of freedom per point in 3 spatial dimensions. For state-of-the-art dwMRI, in which local signal attenuations are recorded under a multitude of magnetic gradient directions, this limitation on angular resolution is too restrictive. The Riemann-Finsler paradigm removes this limitation altogether.

The pivot of Riemann-Finsler geometry is a generalised notion of length of a spatial curve C (*Hilbert's invariant integral* [42]):

$$\mathscr{L}(C) = \int_C F(x, dx)\,. \tag{9}$$

The Lagrangian $F(x, dx)$ is called the *Finsler function*. This function cannot be chosen arbitrarily. In order to interpret Eq. (9) properly as an integral over a one-form, one has to require $F(x, dx) = F(x, \dot{x})dt$ for a parametrized curve $x = x(t)$, with $\dot{x} = dx(t)/dt$, so that the functional $\mathscr{L}(C)$ is well-defined and parameter invariant. More specifically, $F(x, \dot{x})$ is required to be smooth for $\dot{x} \neq 0$ and to satisfy the following properties[5]:

$$F(x, \lambda\dot{x}) = |\lambda| F(x, \dot{x}) \qquad \text{for all } \lambda \in \mathbb{R}, \tag{10}$$

$$F(x, \dot{x}) > 0 \qquad \text{if } \dot{x} \neq 0, \tag{11}$$

$$g_{ij}(x, \dot{x})\xi^i \xi^j > 0 \qquad \text{if } \xi \neq 0, \tag{12}$$

in which the *Riemann-Finsler metric tensor* is defined as

$$g_{ij}(x, \dot{x}) = \frac{1}{2} \frac{\partial^2 F^2(x, \dot{x})}{\partial \dot{x}^i \partial \dot{x}^j}\,. \tag{13}$$

[5]Instead of the norm condition, Eq. (10), one sometimes requires $F(x, \lambda\dot{x}) = \lambda F(x, \dot{x})$. What matters in diffusion processes without convection is *orientation*, not signed direction, so it is natural to require mirror symmetry $\dot{x} \longleftrightarrow -\dot{x}$ a priori.

In these definitions and below, $\dot{x}$ is an a priori *independent* vector argument, not a tangent vector $\dot{x}(t)$ to some underlying parametrized curve $x(t)$, unless explicitly stated otherwise. But it helps intuition to keep in mind the role of this vector argument in an expression like Eq. (9). In particular, when considering a smooth spatial curve $x(t)$, there is a "distinguished" vector $\dot{x} \propto \dot{x}(t)$ associated with any position $x(t)$ along the curve. The extended base manifold with coordinates $(x, \dot{x})$, with $\dot{x} \neq 0$, is referred to as the *slit tangent bundle*. The word "slit" refers to the excluded strip $\dot{x} = 0$. In the context of zero-homogeneous functions, a vector $\dot{x} \neq 0$ represents an equivalence class of points on the line through the origin with direction vector $\dot{x}$. In that case one also refers to the extended base manifold as the *projectivized tangent bundle*, cf. the concept of an *orientation score* by Duits et al. [33–35].

Using Eqs. (10)–(13), it is not difficult to show (with the help of Euler's theorem for homogeneous functions, recall footnote 3) that

$$F^2(x, \dot{x}) = g_{ij}(x, \dot{x})\dot{x}^i \dot{x}^j \,. \tag{14}$$

Riemann's *quadratic restriction* pertains to the "mildly anisotropic" case, $g_{ij}(x, \dot{x}) = g_{ij}(x)$. In general, the Riemann-Finsler metric tensor, Eq. (13), is homogeneous of degree 0: $g_{ij}(x, \lambda\dot{x}) = g_{ij}(x, \dot{x})$ for all $\lambda \in \mathbb{R}$. It may be viewed as being defined on the 5-dimensional projectivized tangent bundle.

Since, in principle, only positions and orientations are of interest, all geometrically relevant quantities will be zero-homogeneous. Although the Finsler function itself does not qualify as such (its domain of definition is the 6-dimensional *slit tangent bundle* of positions and nonzero vectors), it serves as the basic object from which such quantities can be constructed.

The role played by the 3-dimensional (co)tangent spaces erected at each point x of a 3-dimensional Riemannian manifold is replaced by likewise 3-dimensional fibers that collectively constitute a so-called *pulled-back (co)bundle* or *Finsler (co)bundle* in Riemann-Finsler geometry. The major difference is that a pulled-back (co)bundle sits over the 5-dimensional projectivized tangent bundle or 6-dimensional slit tangent bundle, rather than over the 3-dimensional spatial manifold. Given x-coordinates on the spatial manifold the coordinate induced basis sections

$$\left.\frac{\partial}{\partial x^i}\right|_{(x,\dot{x})} \quad \text{respectively} \quad \left. dx^i \right|_{(x,\dot{x})} \tag{15}$$

for its tangent and cotangent bundles can be transplanted to the pulled-back (co)bundle. That is, $\dot{x}$ plays no role in the construction of a fiber at a fiducial point $(x, \dot{x})$.

2.4 Riemann-Finsler Geometry and Its Riemannian Limit

The nontrivial nature of the *Cartan tensor* [42, 43, 48, 60],

$$C_{ijk}(x,\dot{x}) = \frac{1}{4}\frac{\partial^3 F^2(x,\dot{x})}{\partial\dot{x}^i\partial\dot{x}^j\partial\dot{x}^k}\,, \tag{16}$$

distinguishes Riemann-Finsler geometry from its Riemannian counterpart. One can show that $C_{ijk}(x,\dot{x}) = 0$ if and only if space (the x-manifold) is Riemannian. In fact it suffices to inspect the *Cartan one-form*

$$C_i(x,\dot{x}) = g^{jk}(x,\dot{x})C_{ijk}(x,\dot{x})\,, \tag{17}$$

in which the *dual Riemann-Finsler metric tensor*[6] $g^{ij}(x,\dot{x})$ is the inverse of $g_{ij}(x,\dot{x})$:

$$g^{ik}(x,\dot{x})g_{kj}(x,\dot{x}) = \delta^i_j\,. \tag{18}$$

Indeed, one can show that space is Riemannian if and only if the Cartan one-form, Eq. (17), vanishes identically. In view of the significance of zero-homogeneous functions one often encounters the alternative definitions

$$A_{ijk}(x,\dot{x}) = F(x,\dot{x})C_{ijk}(x,\dot{x}) \quad \text{resp.} \quad A_i(x,\dot{x}) = F(x,\dot{x})C_i(x,\dot{x})\,. \tag{19}$$

The dual Riemann-Finsler metric may be used for index raising and lowering, e.g.

$$C^k_{ij}(x,\dot{x}) = g^{k\ell}(x,\dot{x})C_{ij\ell}(x,\dot{x})\,. \tag{20}$$

(There is no ambiguity here by virtue of symmetry of the covariant Cartan tensor.)

Thus the Cartan tensor measures the degree in which the local structure of the Riemann-Finsler manifold deviates from Riemannian. In view of Eqs. (3), (4) (7), and (8) this boils down to a measure for the degree in which the recorded dwMRI data—matched to the basic paradigm, Eq. (2)—violate the validity conditions for DTI. In other words, it provides a (fuzzy) classification of voxels as "DTI-like" (i.e. mildly anisotropic) versus otherwise (i.e. complex or highly anisotropic).

2.5 Connections in Riemann-Finsler Geometry

There is no "obvious" connection (mechanism for parallel transport) on a Riemann-Finsler manifold. The so-called Berwald, Cartan, Chern-Rund and Hashiguchi

[6] It will be seen later, cf. Eqs. (45)–(47), that it is more natural to think of g^{ij} as a metric in q-space, as opposed to the $\dot{x}$-space metric g_{ij}.

connections may all be considered "natural" extensions of the Levi-Civita connection in Riemannian geometry. For instance, the (torsion-free) Chern-Rund connection is defined by[7]

$$\Gamma^i_{jk}(x,\dot{x}) = \frac{1}{2}g^{i\ell}(x,\dot{x})\left(\frac{\delta g_{\ell k}(x,\dot{x})}{\delta x^j} + \frac{\delta g_{j\ell}(x,\dot{x})}{\delta x^k} - \frac{\delta g_{jk}(x,\dot{x})}{\delta x^\ell}\right). \tag{21}$$

This expression is obtained from the "classical" Christoffel symbols of Riemannian geometry by formally replacing the Riemann metric $g_{ij}(x)$ by the Riemann-Finsler metric $g_{ij}(x,\dot{x})$, Eq. (13), and spatial derivatives by the *horizontal vectors*

$$\frac{\delta}{\delta x^i} \stackrel{\text{def}}{=} \frac{\partial}{\partial x^i} - N^j_i(x,\dot{x})\frac{\partial}{\partial \dot{x}^j}. \tag{22}$$

The coefficients $N^j_i(x,\dot{x})$ define the so-called *nonlinear connection* [42]:

$$N^j_i(x,\dot{x}) = \gamma^j_{ik}(x,\dot{x})\dot{x}^k - C^j_{ik}(x,\dot{x})\gamma^k_{\ell m}(x,\dot{x})\dot{x}^\ell\dot{x}^m, \tag{23}$$

in which the *formal Christoffel symbols of the second kind* are introduced as

$$\gamma^i_{jk}(x,\dot{x}) = \frac{1}{2}g^{i\ell}(x,\dot{x})\left(\frac{\partial g_{\ell k}(x,\dot{x})}{\partial x^j} + \frac{\partial g_{j\ell}(x,\dot{x})}{\partial x^k} - \frac{\partial g_{jk}(x,\dot{x})}{\partial x^\ell}\right). \tag{24}$$

Note that in the Riemannian limit, both Eq. (21) as well as Eq. (24) simplify to

$$\Gamma^i_{jk}(x,\dot{x}),\ \gamma^i_{jk}(x,\dot{x}) \longrightarrow \Gamma^i_{jk}(x) = \frac{1}{2}g^{i\ell}(x)\left(\frac{\partial g_{\ell k}(x)}{\partial x^j} + \frac{\partial g_{j\ell}(x)}{\partial x^k} - \frac{\partial g_{jk}(x)}{\partial x^\ell}\right), \tag{25}$$

the standard *Christoffel symbols of the second kind* defining the torsion-free *Levi-Civita connection* in Riemannian geometry. A computation reveals that[8]

$$\begin{aligned}&\Gamma_{ijk}(x,\dot{x}) = \\ &\gamma_{ijk}(x,\dot{x}) - C_{hjk}(x,\dot{x})G^h_{\dot{x}^i}(x,\dot{x}) - C_{hji}(x,\dot{x})G^h_{\dot{x}^k}(x,\dot{x}) + C_{hik}(x,\dot{x})G^h_{\dot{x}^j}(x,\dot{x}),\end{aligned} \tag{26}$$

in which indices have been lowered via the Riemann-Finsler metric tensor:

$$\Gamma_{ijk}(x,\dot{x}) = g_{j\ell}(x,\dot{x})\Gamma^\ell_{ik}(x,\dot{x}) \quad \text{resp.} \quad \gamma_{ijk}(x,\dot{x}) = g_{j\ell}(x,\dot{x})\gamma^\ell_{ik}(x,\dot{x}), \tag{27}$$

[7] Caveat: In [43] Rund defines these symbols as $\Gamma^{*i}_{jk}(x,\dot{x})$.

[8] Caveat: In [43] Rund defines these symbols as $\Gamma^*_{ijk}(x,\dot{x})$.

and in which the *geodesic coefficients* are defined as[9]

$$G^i_{\dot{x}^j}(x,\dot{x}) = \frac{\partial G^i(x,\dot{x})}{\partial \dot{x}^j} \quad \text{with} \quad G^i(x,\dot{x}) = \frac{1}{2}\gamma^i_{jk}(x,\dot{x})\dot{x}^j\dot{x}^k\,. \tag{28}$$

In fact we have

$$G^i_{\dot{x}^j}(x,\dot{x}) = N^i_j(x,\dot{x}) \tag{29}$$

recall Eq. (23).

2.6 Horizontal-Vertical Splitting

The coupling of position and orientation is formalized in terms of the so-called *horizontal* and *vertical basis vectors*, recall Eq. (22),

$$\frac{\delta}{\delta x^i} \stackrel{\text{def}}{=} \frac{\partial}{\partial x^i} - N^\ell_i(x,\dot{x})\frac{\partial}{\partial \dot{x}^\ell} \quad \text{and} \quad \frac{\partial}{\partial \dot{x}^i}\,. \tag{30}$$

These constitute a basis for the *horizontal* and the *vertical tangent bundle over the slit tangent bundle*:

$$\mathrm{H}_{(x,\dot{x})}\mathrm{TM} = \mathrm{span}\left\{\left.\frac{\delta}{\delta x^i}\right|_{(x,\dot{x})}\right\} \quad \text{and} \quad \mathrm{V}_{(x,\dot{x})}\mathrm{TM} = \mathrm{span}\left\{\left.\frac{\partial}{\partial \dot{x}^i}\right|_{(x,\dot{x})}\right\}\,. \tag{31}$$

Their direct sum yields the complete tangent bundle (pointwise):

$$\mathrm{TTM}\backslash\{0\} = \mathrm{HTM} \oplus \mathrm{VTM}\,. \tag{32}$$

By the same token one considers the *horizontal* and *vertical basis covectors*,

$$dx^j \quad \text{and} \quad \delta\dot{x}^i \stackrel{\text{def}}{=} d\dot{x}^i + N^i_\ell(x,\dot{x})dx^\ell\,, \tag{33}$$

yielding the corresponding horizontal and vertical cotangent bundles:

$$\mathrm{H}^*_{(x,\dot{x})}\mathrm{TM} = \mathrm{span}\left\{\left.dx^j\right|_{(x,\dot{x})}\right\} \quad \text{and} \quad \mathrm{V}^*_{(x,\dot{x})}\mathrm{TM} = \mathrm{span}\left\{\left.\delta\dot{x}^i\right|_{(x,\dot{x})}\right\}\,, \tag{34}$$

such that, pointwise,

[9] Caveat: In [42] Bao et al. write $G^i(x,\dot{x}) = \gamma^i_{jk}(x,\dot{x})\dot{x}^j\dot{x}^k$.

$$\mathrm{T}^*\mathrm{TM}\backslash\{0\} = \mathrm{H}^*\mathrm{TM} \oplus \mathrm{V}^*\mathrm{TM}\,. \tag{35}$$

The above vectors and covectors satisfy the following duality relations:

$$dx^i\left(\frac{\delta}{\delta x^j}\right) = \delta\dot{x}^i\left(\frac{\partial}{\partial\dot{x}^j}\right) = \delta^i_j \quad \text{and} \quad dx^j\left(\frac{\partial}{\partial\dot{x}^j}\right) = \delta\dot{x}^i\left(\frac{\delta}{\delta x^j}\right) = 0\,. \tag{36}$$

Incorporating a natural scaling so as to ensure zero-homogeneity with respect to $\dot{x}$ (so that it indeed represents orientation rather than "velocity" or a "displacement") we conclude that

$$\mathrm{TTM}\backslash\{0\} = \mathrm{span}\left\{\frac{\delta}{\delta x^i}\,,\; F(x,\dot{x})\frac{\partial}{\partial\dot{x}^i}\right\}\,, \tag{37}$$

and similarly

$$\mathrm{T}^*\mathrm{TM}\backslash\{0\} = \mathrm{span}\left\{dx^i\,,\; \frac{\delta\dot{x}^i}{F(x,\dot{x})}\right\}\,. \tag{38}$$

The so-called *Sasaki metric* furnishes the slit tangent bundle with a Riemann metric:

$$g(x,\dot{x}) = g_{ij}(x,\dot{x})\,dx^i \otimes dx^j + g_{ij}(x,\dot{x})\,\frac{\delta\dot{x}^i}{F(x,\dot{x})} \otimes \frac{\delta\dot{x}^j}{F(x,\dot{x})}\,. \tag{39}$$

The horizontal and vertical tangent bundles, Eq. (31), are orthogonal relative to this metric.

Cf. the Appendix for the formal motivation of horizontal and vertical basis vectors and covectors. The heuristics behind them is that they permit a coordinate independent, geometrically meaningful splitting into "horizontal" (pertaining to spatial position) and "vertical" (complementary) dimensions. As a counterexample, Eq. (66) in the Appendix shows what happens if we would use the standard coordinate bases in $(x,\dot{x})$-space. A change of spatial coordinates, $x = x(\xi)$, causes the new spatial coordinate vectors to be a linear superposition of *all* coordinate basis vectors that we started out from, whence they do not induce a coordinate independent splitting.

2.7 Horizontal Curves and Finsler Geodesics

Spatial trajectories $x(t)$ have a "natural" (sparse) manifestation in the "vertical" (orientation) dimension, viz. through identification of the trajectory's tangent vector $\dot{x}(t)$ with the vector $\dot{x}$. In other words, interpreted as a curve along the Finsler manifold a spatial curve, $x = \xi(t)$, say, has a natural parametrization $(x,\dot{x}) = (\xi(t),\dot{\xi}(t))$. A tangent vector of this curve is given by (with $\dot{\xi}(t) \equiv d\xi(t)/dt$ and $\ddot{\xi}(t) \equiv d^2\xi(t)/dt^2$)

$$\boldsymbol{T}(t) = \dot{\xi}^i(t)\frac{\partial}{\partial x^i} + \ddot{\xi}^i(t)\frac{\partial}{\partial \dot{x}^i}\,. \tag{40}$$

Note that the individual terms in this equation do not have an intrinsic meaning, to the extent that a splitting of the six dimensional tangent space into span $\{\partial/\partial x^i\}$ and span $\{\partial/\partial \dot{x}^i\}$ is not preserved after a spatial coordinate transformation, cf. Eq. (66) in the Appendix. The aforementioned, geometrically meaningful splitting suggests that we rather decompose the tangent vector as follows, recall Eq. (30):

$$\boldsymbol{T}(t) = \dot{\xi}^i(t)\frac{\delta}{\delta x^i} + \left(\ddot{\xi}^i(t) + N^i_j(\xi(t),\dot{\xi}(t))\,\dot{\xi}^j(t)\right)\frac{\partial}{\partial \dot{x}^i}\,. \tag{41}$$

The requirement of *horizontality* then entails that the vertical component vanishes:

$$\delta\dot{x}^i\,(\boldsymbol{T}(t)) = 0\,. \tag{42}$$

Using the basic duality relations, Eqs. (36), this means that

$$\ddot{\xi}^i(t) + N^i_j(\xi(t),\dot{\xi}(t))\,\dot{\xi}^j(t) = 0\,. \tag{43}$$

By virtue of Eqs. (23) and (24), using the fact that $C^j_{ik}(\xi,\dot{\xi})\dot{\xi}^k = 0$ (a consequence of the homogeneity property $g_{ij}(x,\lambda\dot{x}) = g_{ij}(x,\dot{x})$ and Euler's theorem for homogeneous functions, recall footnote 3), this simplifies to

$$\ddot{\xi}^i(t) + \gamma^i_{jk}(\xi(t),\dot{\xi}(t))\,\dot{\xi}^j(t)\dot{\xi}^k(t) = 0\,. \tag{44}$$

This *geodesic equation* has the same form as in the Riemannian case, except for the fact that Christoffel symbols have been replaced by their formal counterparts, Eq. (24) (or, equivalently, Eq. (21)).

We conclude this section by noting that Eq. (44) provides us with the Finslerian analogue of the geodesic tractography method previously proposed in the Riemannian setting for DTI. We stress that it will likewise need to be complemented with a way to sift geodesics into fibers and non-fibers, either deterministically or probabilistically. The Finslerian analogues of the connectivity measures proposed by Astola et al. [10] are quite straightforward.

2.8 Lagrangian Versus Hamiltonian Frameworks

The non-singular Riemann-Finsler metric enables the same kind of index gymnastics in Riemann-Finsler geometry as the Riemann metric does in the Riemannian case. In particular we have the "velocity"–"momentum" (or $\dot{x}$–q) duality expressed by the equations

$$q_i = g_{ij}(x, \dot{x})\dot{x}^j \quad \text{and} \quad \dot{x}^i = g^{ij}(x, q)q_j \,, \tag{45}$$

in which the *dual Riemann-Finsler metric* has now been prototyped such that

$$g^{ik}(x, q)g_{kj}(x, \dot{x}) = \delta^i_j \,, \tag{46}$$

assuming the aforementioned relationship between $\dot{x}$ and q. Note that, unlike in Eq. (18), the dual metric tensor has been expressed as a function of momentum, not velocity.

The foregoing formulation of the theory, with geometric quantities expressed as functions of position x and velocity $\dot{x}$, is known as the *Lagrangian framework*. The alternative formulation, in which the velocity variable is replaced by momentum q, is known as the *Hamiltonian framework*. The connection between the Lagrangian and corresponding Hamiltonian frameworks is particularly elegant in Riemann-Finsler geometry, in which the *Hamiltonian function* (or *dual Finsler function*) is given by

$$H(x, q) = F(x, \dot{x}) \,, \tag{47}$$

again assuming Eq. (45) to hold. As a consequence, the dual Riemann-Finsler metric tensor plays a similar role in the Hamiltonian framework as the Riemann-Finsler metric tensor does in the Lagrangian framework.[10]

The physical interpretations of the dual formalisms differ and depend on context. The Lagrangian formalism highlights the role of *geodesic congruences*, families of geodesics viewed as "particle trajectories", for which the vector variable $\dot{x}$ expresses "particle velocity". In the Hamiltonian formalism one considers "wave phenomena" induced by such geodesic congruences, in which case the covector variable q enters as "wave momentum", which, by definition, is the normal along which wave fronts propagate. Recall that in anisotropic media wave fronts induced by the interference of the disturbances caused by individual particles do not travel in the same direction as the particles themselves (cf. Huygens' principle, [61]). This is expressed by Eq. (45), as the (dual) metric is not necessarily diagonal.

2.9 Indicatrix and Figuratrix

The *indicatrix* at a fixed point x is the level set, or "glyph", of the Riemann-Finsler unit sphere, $F(x, \dot{x}) = 1$, or, by virtue of Eq. (14),

$$g_{ij}(x, \dot{x})\dot{x}^i\dot{x}^j = 1 \,. \tag{48}$$

[10] One sometimes reserves the terms Lagrangian and Hamiltonian for the *squared* Finsler and dual Finsler function. The associated "energy" functionals are *not* parametrization invariant.

The *figuratrix* at a fixed point x is the Hamiltonian counterpart, i.e. the level set given by $H(x,q) = 1$, recall Eqs. (4) and (8):

$$g^{ij}(x,q)q_i q_j = 1\,. \tag{49}$$

One can show that, as a result of zero-homogeneity of the Riemann-Finsler (dual) metric tensor, both indicatrix as well as figuratrix represent *convex glyphs*.

A convenient interpretation of these structures is obtained by "freezing" the (co)vector argument of the Riemann-Finsler (dual) metric tensor in Eqs. (48) and (49), so that one ends up with (parametrized) quadratic forms. These are known as the *osculating indicatrix* and *osculating figuratrix*, respectively:

$$g_{ij}(x,\dot{x}_0)\dot{x}^i\dot{x}^j = 1\,, \tag{50}$$

$$g^{ij}(x,q_0)q_i q_j = 1\,. \tag{51}$$

One could think of these as gauge figures of a parametrized family of inner products on the tangent, respectively cotangent space of the spatial domain, each direction (specified by $\dot{x}_0$ or q_0) having its own unique instance. The Cartan tensor, Eq. (16), plays the pivotal role in relating the individual members of such a family.

In the DTI/Riemannian case the coefficients in Eqs. (50) and (51) are independent of the orientation parameters, so that each point in space has an unambiguously defined ellipsoidal shape representing the entire family. Indicatrices have been widely adopted in DTI visualization [62]. They might also be useful for our general case, although by their convex nature they are not likely to reflect the rich amount of information contained in a general Finsler function very clearly. A slick selection of osculating indicatrices might in that case prove more insightful.

2.10 Covariant Derivatives

The horizontal and vertical one-forms given by Eq. (38) can be used as a basis for decomposing the *covariant differential* of an arbitrary tensor field on the slit tangent bundle. For simplicity consider

$$\boldsymbol{T}(x,\dot{x}) = T^i_j(x,\dot{x})\frac{\partial}{\partial x^i}\otimes dx^j\,, \tag{52}$$

and

$$\nabla\boldsymbol{T}(x,\dot{x}) = (\nabla\boldsymbol{T})^i_j\,(x,\dot{x})\frac{\partial}{\partial x^i}\otimes dx^j\,. \tag{53}$$

Then each component on the r.h.s. is a one-form, and can thus be written as a sum of horizontal and vertical one-forms. By definition,

$$(\nabla \boldsymbol{T})^i_j\,(x,\dot{x}) = T^i_{j|k}(x,\dot{x})\,dx^k + T^i_{j;k}(x,\dot{x})\,\frac{\delta \dot{x}^k}{F(x,\dot{x})}\,. \tag{54}$$

By evaluation on the corresponding dual basis, Eq. (37), one obtains

$$T^i_{j|k}(x,\dot{x}) = \frac{\delta T^i_j(x,\dot{x})}{\delta x^k} + T^\ell_j(x,\dot{x})\Gamma^i_{\ell k}(x,\dot{x}) - T^i_\ell(x,\dot{x})\Gamma^\ell_{jk}(x,\dot{x})\,, \tag{55}$$

$$T^i_{j;k}(x,\dot{x}) = F(x,\dot{x})\,\frac{\partial T^i_j(x,\dot{x})}{\partial \dot{x}^k}\,. \tag{56}$$

Equations (55) and (56) are the components of the *horizontal covariant derivative* and the *vertical covariant derivative* of the tensor field, respectively (relative to the Chern-Rund connection, recall Eq. (21)). Higher order tensors are treated similarly. Their horizontal covariant derivatives will contain as many "correction terms" involving the Riemann-Finsler Γ-symbols of Eq. (21) as indicated by their order. Note the elegant similarity with the Riemannian case.

Some cases are particularly important, e.g. those involving the Riemann-Finsler metric tensor or its dual. We have

$$g_{ij|k}(x,\dot{x}) = 0\,, \tag{57}$$

$$g_{ij;k}(x,\dot{x}) = 2F(x,\dot{x})\,C_{ijk}(x,\dot{x})\,, \tag{58}$$

$$g^{ij}{}_{|k}(x,\dot{x}) = 0\,, \tag{59}$$

$$g^{ij}{}_{;k}(x,\dot{x}) = -2F(x,\dot{x})\,C^{ij}_k(x,\dot{x})\,. \tag{60}$$

The Kronecker tensor is covariantly constant both horizontally as well as vertically:

$$\delta^i_{j\,|k} = 0\,, \tag{61}$$

$$\delta^i_{j;k} = 0\,. \tag{62}$$

Thus, unlike in the Riemannian case, the Riemann-Finsler metric tensor is covariantly constant only along horizontal directions, whereas its behavior in vertical directions is governed by the Cartan tensor (the covariant derivative is said to be "almost metric compatible").

3 Conclusion and Discussion

Riemann-Finsler geometry naturally extends the Riemannian rationale used in the context of DTI to general dwMRI representations. It can be equivalently approached from a Lagrangian or Hamiltonian perspective, although the latter appears to be most closely related to the physics of dwMRI acquisition and its underlying model in terms of a generalized mono-exponential Stejskal-Tanner equation.

We have pointed out its potential application to voxel classification based on the Cartan tensor and related quantities, and to dwMRI tractography by deriving the corresponding Finsler geodesic equations, without the quadratic restriction inherent to the DTI model, yet retaining quadratic scaling in the *magnitude* of the gradient magnetic field. Although this does not cover the general (multi-exponential and/or non-homogeneous) case, the conditions for and limitations of this conjecture, and in particular the added value relative to DTI, diffusional kurtosis imaging (DKI), and (other) HARDI schemes, are worthwhile investigating. Future work will concentrate on this, on the reconstruction of the (dual) Finsler function and related quantities, and on experimental validation of Finsler tractography and voxel classification as advocated in this chapter.

Appendix: Horizontal and Vertical Splitting

We may consider the partial derivatives with respect to x^i and $\dot{x}^i$ as coordinate vector fields on the tangent bundle TM, and consider the effect induced by a change of coordinates of the base manifold M, $x = x(\xi)$ say. Since $\dot{x}$ is a vector, this induces the following *vector transformation law* for its components $\dot{x}^i$ expressed in terms of its new components, $\dot{\xi}^p$, say:

$$\dot{x}^i = \frac{\partial x^i}{\partial \xi^p}\dot{\xi}^p \,, \tag{63}$$

or, equivalently,

$$\frac{\partial}{\partial \dot{\xi}^p} = \frac{\partial x^i}{\partial \xi^p}\frac{\partial}{\partial \dot{x}^i} \,, \tag{64}$$

so that, by construction,

$$\dot{x}^i \frac{\partial}{\partial x^i} = \dot{\xi}^p \frac{\partial}{\partial \xi^p} \,. \tag{65}$$

As a result,

$$\frac{\partial}{\partial \xi^p} = \frac{\partial x^i}{\partial \xi^p}\frac{\partial}{\partial x^i} + \frac{\partial^2 x^i}{\partial \xi^p \partial \xi^q}\dot{\xi}^q \frac{\partial}{\partial \dot{x}^i} \,. \tag{66}$$

Given the definition of the horizontal vectors, Eq. (22), and of the nonlinear connection, Eq. (23), it is then a tedious but straightforward exercise to deduce that

$$\frac{\delta}{\delta \xi^p} = \frac{\partial x^i}{\partial \xi^p}\frac{\delta}{\delta x^i} \,, \tag{67}$$

similar to the vector transformation law for the vertical components, recall Eq. (64).

Likewise one has the *covector transformation law* for the components of the horizontal and vertical one-forms, recall Eq. (33):

$$dx^i = \frac{\partial x^i}{\partial \xi^p} d\xi^p \,, \tag{68}$$

respectively

$$\delta \dot{x}^i = \frac{\partial x^i}{\partial \xi^p} \delta \dot{\xi}^i \,. \tag{69}$$

The "natural" transformation behavior expressed by Eqs. (64) and (67)–(69) motivates the definitions of horizontal and vertical vectors and covectors in Sect. 2.6.

References

1. Hagmann, P., Jonasson, L., Maeder, P., Thiran, J.P., Wedeen, V.J., Meuli, R.: Understanding diffusion MR imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond. RadioGraphics **26**, S205 (2006)
2. Basser, P.J., Mattiello, J., Le Bihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson. **103**, 247 (1994)
3. Basser, P.J., Mattiello, J., Le Bihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**(1), 259 (1994)
4. Basser, P.J., Mattiello, J., Le Bihan, D.: Diffusion tensor MR imaging of the human brain. Radiology **201**(3), 637 (1996)
5. Basser, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J. Magn. Reson. Ser. B **111**(3), 209 (1996)
6. Basser, P.J.: Relationships between diffusion tensor and q-space MRI. Magn. Reson. Med. **47**(2), 392 (2002)
7. Le Bihan, D., Mangin, J.F., Poupon, C., Clark, C.A., Pappata, S., Molko, N., Chabriat, H.: Diffusion tensor imaging: concepts and applications. J. Magn. Reson. Imaging **13**, 534 (2001)
8. Moseley, M., Cohen, Y., Kucharczyk, J., Mintorovitch, J., Asgari, H.S., Wendland, M.F., Tsuruda, J., Norman, D.: Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. Radiology **176**(2), 439 (1990)
9. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magn. Reson. Med. **56**(2), 411 (2006)
10. Astola, L., Florack, L., ter Haar Romeny, B.: Measures for pathway analysis in brain white matter using diffusion tensor images. In: Karssemeijer, N., Lelieveldt, B. (eds.) Proceedings of the Twentieth International Conference on Information Processing in Medical Imaging–IPMI 2007, Kerkrade. Lecture Notes in Computer Science, vol. 4584, pp. 642–649. Springer, Berlin (2007)
11. Astola, L., Florack, L.: Sticky vector fields and other geometric measures on diffusion tensor images. In: Proceedings of the 9th IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis, held in conjuction with the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, 23–28 June 2008. IEEE Computer Society
12. Astola, L., Fuster, A., Florack, L.: A Riemannian scalar measure for diffusion tensor images. Pattern Recognit. **44**(9), 1885 (2011). doi:10.1016/j.patcog.2010.09.009

13. Deriche, R., Calder, J., Descoteaux, M.: Optimal real-time Q-ball imaging using regularized Kalman filtering with incremental orientation sets. Med. Image Anal. **13**(4), 564 (2009)
14. Fillard, P., Pennec, X., Arsigny, V., Ayache, N.: Clinical DT-MRI estimation, smoothing, and fiber tracking with log-Euclidean metrics. IEEE Trans. Med. Imaging **26**(11) (2007)
15. Florack, L.M.J., Astola, L.J.: A multi-resolution framework for diffusion tensor images. In: Aja Fernández, S., de Luis Garcia, R. (eds.) CVPR Workshop on Tensors in Image Processing and Computer Vision, Anchorage, 24–26 June 2008. IEEE (2008). Digital proceedings
16. Lenglet, C., Deriche, R., Faugeras, O.: Inferring white matter geometry from diffusion tensor MRI: application to connectivity mapping. In: Pajdla, T., Matas, J. (eds.) Proceedings of the Eighth European Conference on Computer Vision, Prague, May 2004. Lecture Notes in Computer Science, vol. 3021–3024, pp. 127–140. Springer, Berlin (2004)
17. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing. J. Math. Imaging Vis. **25**(3), 423 (2006)
18. Lenglet, C., Prados, E., Pons, J.P.: Brain connectivity mapping using Riemannian geometry, control theory and PDEs. SIAM J. Imaging Sci. **2**(2), 285 (2009)
19. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. Int. J. Comput. Vis. **66**(1), 41 (2006)
20. Prados, E., Soatto, S., Lenglet, C., Pons, J.P., Wotawa, N., Deriche, R., Faugeras, O.: Control theory and fast marching techniques for brain connectivity mapping. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, June 2006, vol. 1, pp. 1076–1083. IEEE Computer Society (2006)
21. Florack, L., Balmashnova, E.: Two canonical representations for regularized high angular resolution diffusion imaging. In: Alexander, D., Gee, J., Whitaker, R. (eds.) MICCAI Workshop on Computational Diffusion MRI, New York, 10 Sept 2008, pp. 85–96 (2008)
22. Florack, L., Balmashnova, E., Astola, L., Brunenberg, E.: A new tensorial framework for single-shell high angular resolution diffusion imaging. J. Math. Imaging Vis. **3**(38), 171 (2010). Published online: doi:10.1007/s10851-010-0217-3
23. Jensen, J.H., Helpern, J.A., Ramani, A., Lu, H., Kaczynski, K.: Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. Magn. Reson. Med. **53**(6), 1432 (2005)
24. Jian, B., Vemuri, B.C., Özarslan, E., Carney, P.R., Mareci, T.H.: A novel tensor distribution model for the diffusion-weighted MR signal. NeuroImage **37**, 164 (2007)
25. Liu, C., Bammer, R., Acar, B., Moseley, M.E.: Characterizing non-Gaussian diffusion by using generalized diffusion tensors. Magn. Reson. Med. **51**(5), 924 (2004)
26. Özarslan, E., Mareci, T.H.: Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution imaging. Magn. Reson. Med. **50**(5), 955 (2003)
27. Özarslan, E., Shepherd, T.M., Vemuri, B.C., Blackband, S.J., Mareci, T.H.: Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). NeuroImage **31**, 1086 (2006)
28. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Apparent diffusion coefficients from high angular resolution diffusion imaging: estimation and applications. Magn. Reson. Med. **56**(2), 395 (2006)
29. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical Q-ball imaging. Magn. Reson. Med. **58**(3), 497 (2007)
30. Florack, L.M.J.: Codomain scale space and regularization for high angular resolution diffusion imaging. In: Aja Fernández, S., de Luis Garcia, R. (eds.) CVPR Workshop on Tensors in Image Processing and Computer Vision, Anchorage, 24–26 June 2008. IEEE (2008). Digital proceedings
31. Hess, C.P., Mukherjee, P., Tan, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. Magn. Reson. Med. **56**, 104 (2006)
32. Tuch, D.S.: Q-ball imaging. Magn. Reson. Med. **52**, 1358 (2004)

33. Duits, R., Franken, E.M.: Left invariant parabolic evolution equations on SE(2) and contour enhancement via invertible orientation scores, part I: linear left-invariant diffusion equations on SE(2). Q. Appl. Math. **68**(2), 255 (2010)
34. Duits, R., Franken, E.M.: Left invariant parabolic evolution equations on SE(2) and contour enhancement via invertible orientation scores, part II: nonlinear left-invariant diffusion equations on invertible orientation scores. Q. Appl. Math. **68**(2), 293 (2010)
35. Duits, R., Franken, E.: Left-invariant diffusions on the space of positions and orientations and their application to crossing-preserving smoothing of HARDI images. Int. J. Comput. Vis. **12**(3), 231 (2011). Published online: doi:10.1007/s11263-010-0332-z
36. O'Donnell, L., Haker, S., Westin, C.F.: New approaches to estimation of white matter connectivity in diffusion tensor MRI: elliptic PDEs and geodesics in a tensor-warped space. In: Dohi, T., Kikinis, R. (eds.) Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002, Tokyo, 25–28 Sept 2002. Lecture Notes in Computer Science, vol. 2488–2489, pp. 459–466. Springer, Berlin (2002)
37. Astola, L.J., Florack, L.M.J.: Finsler geometry on higher order tensor fields and applications to high angular resolution diffusion imaging. In: Tai, X.C., Mørken, K., Lysaker, M., Lie, K.A. (eds.) Scale Space and Variational Methods in Computer Vision: Proceedings of the Second International Conference, SSVM 2009, Voss. Lecture Notes in Computer Science, vol. 5567, pp. 224–234. Springer, Berlin (2009)
38. Astola, L.J.: Multi-scale Riemann-Finsler geometry: applications to diffusion tensor imaging and high resolution diffusion imaging. Ph.D. thesis, Eindhoven University of Technology, Department of Mathematics and Computer Science, Eindhoven (2010)
39. Astola, L.J., Florack, L.M.J.: Finsler geometry on higher order tensor fields and applications to high angular resolution diffusion imaging. Int. J. Comput. Vis. **92**(3), 325 (2011). doi:10.1007/s11263-010-0377-z
40. Astola, L.J., Jalba, A.C., Balmashnova, E.G., Florack, L.M.J.: Finsler streamline tracking with single tensor orientation distribution function for high angular resolution diffusion imaging. J. Math. Imaging Vis. **41**(3), 170 (2011)
41. Melonakos, J., Pichon, E., Angenent, S., Tannenbaum, A.: Finsler active contours. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 412 (2008)
42. Bao, D., Chern, S.S., Shen, Z.: An Introduction to Riemann-Finsler Geometry. Graduate Texts in Mathematics, vol. 2000. Springer, New York (2000)
43. Rund, H.: The Differential Geometry of Finsler Spaces. Springer, Berlin (1959)
44. Shen, Z.: Lectures on Finsler Geometry. World Scientific, Singapore (2001)
45. Mo, X.: An Introduction to Finsler Geometry. Peking University Series in Mathematics, vol. 1. World Scientific, Singapore (2006)
46. Riemann, B.: Über die Hypothesen, welche der Geometrie zu Grunde liegen. In: Weber, H. (ed.) Gesammelte Mathematische Werke, pp. 272–287. Teubner, Leibzig (1892)
47. Finsler, P.: Ueber kurven und Flächen in allgemeinen Räumen. Ph.D. thesis, University of Göttingen,, Göttingen (1918)
48. Cartan, E.: Les Espaces de Finsler. Hermann, Paris (1934)
49. Michailovich, O., Rathi, Y., Dolui, S.: Spatially regularized compressed sensing for high angular resolution diffusion imaging. IEEE Trans. Med. Imaging **30**(5), 1100 (2011). doi:10.1109/TMI.2011.2142189
50. Stejskal, E.O., Tanner, J.E.: Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. J. Comput. Phys. **42**(1), 288 (1965)
51. Bloch, F.: Nuclear induction. Phys. Rev. **70**, 460 (1946)
52. Haacke, E.M., Brown, R.W., Thompson, M.R., Venkatesan, R.: Magnetic Resonance Imaging: Physical Principles and Sequence Design. Wiley, New York (1999)
53. Stejskal, E.O.: Use of spin echoes in a pulsed magnetic-field gradient to study anisotropic, restricted diffusion and flow. J. Comput. Phys. **43**(10), 3597 (1965)
54. Torrey, H.C.: Bloch equations with diffusion terms. Phys. Rev. D **104**, 563 (1956)

55. Sinnaeve, D.: The Stejskal-Tanner equation generalized for any gradient shape—an overview of most pulse sequences measuring free diffusion. Concepts Magn. Reson. Part A **40A**(2), 39 (2012). doi:10.1002/cmr.a.21223
56. Panagiotaki, E., Schneider, T., Siow, B., Hall, M.G., Lythgoe, M.F., Alexander, D.C.: Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. NeuroImage **59**(3), 2241 (2012). doi:10.1016/j.neuroimage.2011.09.081
57. Fieremans, E., Jensen, J.H., Helpern, J.A.: White matter characterization with diffusional kurtosis imaging. NeuroImage **58**(1), 177 (2011). doi:10.1016/j.neuroimage.2011.06.006
58. Hao, X., Whitaker, R.T., Fletcher, P.T.: Adaptive Riemannian metrics for improved geodesic tracking of white matter. In: Székely, G., Hahn, H.K. (eds.) Proceedings of the Twenty-Second International Conference on Information Processing in Medical Imaging–IPMI 2011, Kloster Irsee. Lecture Notes in Computer Science, vol. 6801, pp. 13–24. Springer, Berlin (2011)
59. Fuster, A., Dela Haije, T.C.J., Florack, L.M.J.: On the Riemannian rationale for diffusion tensor imaging. In: Y. Wiaux, E.D. McEwen (eds.) Proceedings of the International BASP Frontiers Workshop 2013, BASP 2013, Villars-Sur-Ollon, 27 Jan–1 Feb 2013, p. 62 (2013) http://baspfrontiers.epfl.ch/proceedings.pdf
60. Rund, H.: The Hamilton-Jacobi Theory in the Calculus of Variations. Robert E. Krieger Publishing Company, Huntington (1973)
61. Huygens, C.: Traité de la Lumière. Pierre van der Aa, Leiden (1690)
62. Kingsley, P.B.: Introduction to diffusion tensor imaging mathematics: part I. Tensors, rotations, and eigenvectors. Concepts Magn. Reson. Part A **28A**(2), 101 (2006). doi:10.1002/cmr.a.20048

Riemann-Finsler Multi-valued Geodesic Tractography for HARDI

Neda Sepasian, Jan H.M. ten Thije Boonkkamp, Luc M.J. Florack, Bart M. Ter Haar Romeny, and Anna Vilanova

Abstract We introduce a geodesic based tractography method for High Angular Resolution Diffusion Imaging (HARDI). The concepts used are similar to the ones in geodesic based tractography for Diffusion Tensor Imaging (DTI). In DTI, the inverse of the second-order diffusion tensor is used to define the manifold where the geodesics are traced. HARDI models have been developed to resolve complex fiber populations within a voxel, and higher order tensors (HOT) are possible representations for HARDI data. In our framework, we apply Finsler geometry, which extends Riemannian geometry to a directionally dependent metric. A Finsler metric is defined in terms of HARDI higher order tensors. Furthermore, the Euler-Lagrange geodesic equations are derived based on the Finsler geometry. In contrast to other geodesic based tractography algorithms, the multi-valued numerical solution of the geodesic equations can be obtained. This gives the possibility to capture all geodesics arriving at a single voxel instead of only computing the shortest one. Results are analyzed to show the potential and characteristics of our algorithm.

N. Sepasian • L.M.J. Florack
Department of Biomedical Engineering, Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: nsepasian@gmail.com; L.M.J.Florack@tue.nl

J.H.M. ten Thije Boonkkamp
Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: J.H.M.tenThijeBoonkkamp@tue.nl

B.M. Ter Haar Romeny
Department of Biomedical Engineering, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands
e-mail: B.M.terHaarRomeny@tue.nl

A. Vilanova (✉)
Faculty Electrical Engineering, Mathematics and Computer Science, Computer Graphics and Visualization, Delft University of Technology, Delft, The Netherlands
e-mail: A.Vilanova@tudelft.nl

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_9,

1 Introduction

Diffusion-Weighted Magnetic Resonance Imaging(DW-MRI) measures water diffusion characteristics in tissue for a given direction. The diffusion profile in a specific location can be obtained by combining the DW-MRI measurements in different directions. The diffusion profile gives information about the underlying fibrous structure, e.g., in human brain white matter, based on the assumption that water molecules are moving less freely perpendicular to the fibrous structure than along the fiber tracts. Techniques to reconstruct fiber tracts based on the diffusion profiles are known as tractography or fiber tracking methods.

From DW-MRI measurements often a positive definite second-order tensor is defined, referred to as diffusion tensor imaging (DTI) [4]. Despite the simplicity of the model, this technique is shown to be promising to reveal the structure of brain white matter. However, DTI assumes that each voxel contains fibers with only one orientation, and it is known that in white matter often multiple fiber orientations occur [2]. High Angular Resolution Diffusion Imaging (HARDI) and its modeling techniques have been developed to overcome the limitations of the DTI model [12, 19, 27, 29–31].

The models applied to HARDI data result in a function on the unit sphere that gives information about the diffusion profile within the voxel. This function on the unit sphere is obtained by different assumptions and models of the diffusion and acquisition process. In general, the diffusion profiles are assumed to have local maxima in the orientations of the underlying fiber tracts. One of the most popular models is to use the Orientation Distribution Function (ODF) [12, 28, 30]. An extensive description of the different HARDI models is considered outside the scope of this chapter. However, it should be noticed that any function on the unit sphere, and therefore any HARDI model, can be represented by Higher Order Tensors (HOT).

Numerous tractography algorithms have been introduced to reconstruct the fibrous structure from DTI and HARDI data. In the most commonly used tractography algorithms, i.e., streamline based methods, the fibers are estimated by using a number of directions (i.e., the principal direction of the diffusion tensor [4] or the local maxima in the HARDI models [15]). These methods are based on local characteristics and therefore sensitive to noise. A possible solution to resolve these limitations of classic tractography, is to apply global approaches such as geodesic based algorithms [7, 14, 16]. These techniques are based on the assumption that fibers follow the most efficient diffusion propagation paths. A Riemannian manifold is defined using as metric the inverse of the diffusion tensor. Paths in this manifold are shorter if the diffusion is stronger along that path. Therefore, geodesics (i.e., shortest paths) in this manifold follow the most efficient diffusion paths. The global approaches are considerably more computationally expansive than the streamline based methods.

Sepasian et al. [24] presented a ray-tracing algorithm for computing geodesics in anisotropic domains. Common geodesic based methods obtain the viscosity solution based on the Hamilton-Jacobi equation and by solving it numerically using

different numerical schemes, e.g., fast-sweeping. These methods are robust and mathematically elegant. However, the viscosity solution provides only one geodesic between two given points in the domain. In contrast, the approach of Sepasian et al. [24] can capture multi-valued geodesics connecting two given points by considering the geodesics as function of position and direction. Moreover, it is based on the Euler-Lagrange (EL) equations, and therefore local changes in the geodesic can be taken into account. However, this method remains limited to Riemannian manifolds and, therefore, to DTI models.

Campbell [8] proposed a front evolution approach based on HARDI. Pichon et al. [22] introduced a variational cost function depending on all directions on the unit sphere. This directional information is obtained using HARDI. Péchaud et al. [21] presented an algorithm for the calculation of shortest paths on a manifold defined by ODFs. The metric for each position is defined as the orientation distribution function and the geodesics locally follow the paths going through areas of high diffusion.

The methods mentioned above belong to the class of deterministic tractography methods, i.e., given the same input these methods will always give the same result. In the generic case, geodesic-based methods will find a finite number of geodesic paths given two points in the domain. Probabilistic tractography constitutes another class of methods where the variation of the pathways due to model assumptions and/or noise is considered. A probability distribution is built, and based on this distribution, a random process generates many paths originating from one initial position [10, 20]. HARDI tractography techniques are often probabilistic [5, 13].

In this chapter, we propose an extension to HARDI data of the method introduced by Sepasian et al. [24]. Riemannian metric depends on the position in space, and it has no directional dependency. A Riemannian framework is well suited for DTI data, but not for HARDI data. Finsler geometry is the natural extension of the Riemannian geometry to directional dependent metrics. In a Finsler metric, for each position and direction we might have a different metric tensor. Therefore, to extend the work of Sepasian et al. [24] to HARDI data, we need to extend it to Finsler geometry. You can find further explanation on the Riemannian-Finsler geometry in the chapter *Riemann-Finsler Geometry for Diffusion Weighted Magnetic Resonance Imaging*. In comparison to previous work, e.g., Péchaud et al. [21], our technique is based on minimizing the length using the geodesic equations in Finsler geometry instead of a general cost function. Melonakos et al. [17] presented a tractography method for Finsler geometry within their active contours segmentation framework. Their method is based on DTI data and the Finsler metric is used to reduce signal to noise ratio. Furthermore, we uniquely compute the multi-valued solution of the geodesic equations between two points in the domain instead of the single-valued viscosity solution.

The chapter is organized as follows. In Sect. 2 we introduce Finsler geometry which extends Riemannian geometry to directionally dependent metrics. We introduce the Euler-Lagrange form of the geodesic equations in the Finsler geometry in Sect. 3. Next, in Sect. 4, we describe the numerical model necessary for the implementation of the presented ray-tracing algorithm. Finally, preliminary results are presented in Sect. 5.

2 Finsler Metric

In this chapter, we consider the generalization to higher order tensors as an extension of the second order tensors. Higher order tensors allow the representation of multiple fiber orientations. For simplicity, we present the theory for fourth order tensors, although, its extension to higher orders can be trivially deduced. $P_4(\mathbf{x}, \mathbf{y})$ is a function on the unit sphere that represent the diffusion profile obtained by one of the HARDI modeling techniques. At this point, the theory is presented in a general form such that $P_4(\mathbf{x}, \mathbf{y})$ can be the diffusion profile resulting from diverse HARDI modeling techniques, as long as the diffusion flux increases with increasing values of $P_4(\mathbf{x}, \mathbf{y})$.

We fit the fourth order tensor coefficients $\mathbf{D} = \mathbf{D}(\mathbf{x})$ to the function $P_4(\mathbf{x}, \mathbf{y})$ by using the sampled data on the unit sphere [12], i.e.,

$$P_4(\mathbf{x}, \mathbf{y}) = D_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x}) y^{\alpha_1} y^{\alpha_2} y^{\alpha_3} y^{\alpha_4}, \tag{1}$$

with $\alpha_i = 1, 2, 3$, $i = 1, 2, 3, 4$, $\mathbf{x}$ contains the spatial coordinates and $\mathbf{y} = (y^{\alpha_i}) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ is the directional vector with $\phi \in [0, 2\pi)$ and $\theta \in [0, \pi]$. Here, the coefficients $D_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x})$ are the elements of $\mathbf{D}$. Note that in all the formulas the Einstein summation convention is used, i.e., we sum over repeated indices, one in a superscript and one in subscript position. The tensor $\mathbf{D}$ satisfies the symmetry property,

$$D_{\alpha_1\alpha_2\alpha_3\alpha_4} = D_{\sigma(\alpha_1)\sigma(\alpha_2)\sigma(\alpha_3)\sigma(\alpha_4)}, \tag{2}$$

for any permutation σ. Therefore we can reduce the number of components from 81 to 15.

Similar to the DTI Riemannian framework, we use the heuristic that a high probability of finding a fiber in direction $\mathbf{y}$ corresponds to a larger diffusivity represented in $P_4(\mathbf{x}, \mathbf{y})$, and a shorter travel time for the diffusing particle. Therefore, the metric should give the shortest distance in the direction where diffusion is largest. In DTI this is achieved by introducing the metric as the inverse of the diffusion tensor. Consequently, the largest eigenvalue of the diffusion tensor becomes the smallest one for the metric. In contrast, in HARDI we deal with more complex diffusion profiles. We need to extend the framework to a Finsler geometry where the metric is function of position and direction. Furthermore, we need to find a suitable framework for inverting the HOT. A proper inversion should preserve certain properties including the average value of the function and the angle between two directions [6]. Astola et al. [3] suggest to use spherical inversion for the HARDI diffusion profile. Figure 1 illustrates the spherical inversion of a point on a surface M with coordinates $\mathbf{x}_M$. The inverse of a point M with respect to a reference sphere centered at the origin O with radius $r = a$ is the point M' such that $|\mathbf{x}_M| \times |\mathbf{x}_{M'}| = a^2$ where $\mathbf{x} = r\mathbf{e}_r$ is the position vector and $\mathbf{e}_r$ is the radial unit vector in spherical coordinates. The points M and M' are on the same ray through O. Figure 1 shows that the inversion maps points that were outside the

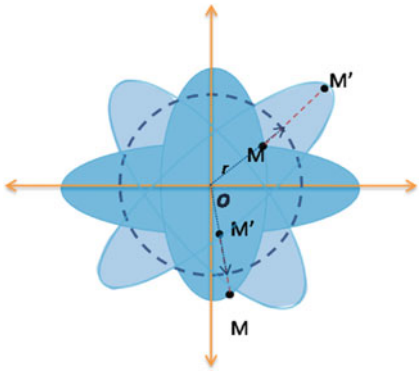


Fig. 1 An illustration of spherical inversion in two dimensions. M' is the inverse of M with respect to the circle

sphere to points that are inside the sphere, and vice versa. For example, for the case when M is outside the reference sphere, $|\mathbf{x}_M| > a$ which gives

$$|\mathbf{x}_{M'}| = \frac{a^2}{|\mathbf{x}_M|} < a.$$

In analogy with the spherical inversion, we define $\tilde{P}_4(\mathbf{x}, \mathbf{y})$, which is the inverse of $P_4(\mathbf{x}, \mathbf{y})$, as follows

$$\tilde{P}_4(\mathbf{x}, \mathbf{y}) = \frac{\overline{P_4}(\mathbf{x})}{P_4(\mathbf{x}, \mathbf{y})} = \tilde{D}_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x}) y^{\alpha_1} y^{\alpha_2} y^{\alpha_3} y^{\alpha_4}, \tag{3}$$

where $\tilde{\mathbf{D}}$ is the HOT that fits $\tilde{P}_4(\mathbf{x}, \mathbf{y})$ and $\overline{P_4}(\mathbf{x})$ is the average of the HOT over the unit sphere, i.e.,

$$\overline{P_4}(\mathbf{x}) = \int_{|\mathbf{y}|=1} P_4(\mathbf{x}, \mathbf{y}) \mathrm{d}\mathbf{y}.$$

Astola et al. [3] propose the following Finsler norm for the fourth order tensors

$$F(\mathbf{x}, \mathbf{y}) = (\tilde{P}_4(\mathbf{x}, \mathbf{y}))^{1/4}. \tag{4}$$

It can be shown that the Finsler metric is given by the bilinear form $F = F(\mathbf{x}, \mathbf{y})$ as follows

$$g_{\alpha\beta}(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \frac{\partial^2 F^2}{\partial y^\alpha \partial y^\beta} \tag{5}$$

where α and β are used to index the components of the tensor g.

In the following, we illustrate that the Riemannian metric is a special case of the Finsler metric. If $F^2(\mathbf{x}, \mathbf{y}) = g_{\alpha\beta}(\mathbf{x}) y^\alpha y^\beta$, i.e., $g_{\alpha\beta}$ only depends on $\mathbf{x}$, then

Fig. 2 Illustration of the directional dependency of tensors $g_{\alpha\beta}(\mathbf{x},\mathbf{y})$. The $\tilde{P}_4(\mathbf{x},\mathbf{y})$ profile is shown as a surface in all images. The *arrows* indicate specific directions $\mathbf{y}$, and the ellipsoids represent the tensors $g_{\alpha\beta}(\mathbf{x},\mathbf{y})$ calculated from $\tilde{P}_4(\mathbf{x},\mathbf{y})$ by fixing the direction $\mathbf{y}$ for each image

$$g_{\alpha\beta}(\mathbf{x},\mathbf{y}) = \frac{1}{2}\frac{\partial^2}{\partial y_\alpha \partial y_\beta}\left(g_{\alpha\beta}(\mathbf{x})y^\alpha y^\beta\right) = g_{\alpha\beta}(\mathbf{x}), \tag{6}$$

known as the Riemannian metric. In contrast, Finsler metric does not only depend on the location $\mathbf{x}$ but also on the direction $\mathbf{y}$.

The necessary conditions of differentiability, homogeneity and strong convexity for (4) have been studied in Astola et al. [3] and the strong convexity criterion only holds if the second-order tensor $D_{\alpha_1\alpha_2\alpha_3\alpha_4}y^{\alpha_1}y^{\alpha_2}$ is positive definite for every $\mathbf{y}$. In practice this condition turns out to be satisfied.

Substituting (4) in the bilinear form (5), we can show that the Finsler metric tensor reads

$$\begin{aligned} g_{\alpha\beta}(\mathbf{x},\mathbf{y}) &= \frac{1}{2}\frac{\partial^2}{\partial y^\alpha y^\beta}F^2(\mathbf{x},\mathbf{y}) \\ &= -2\tilde{P}_4(\mathbf{x},\mathbf{y})^{-3/2}\left(\tilde{D}_{\alpha\alpha_1\alpha_2\alpha_3}y^{\alpha_1}y^{\alpha_2}y^{\alpha_3}\right)\left(\tilde{D}_{\beta\beta_1\beta_2\beta_3}y^{\beta_1}y^{\beta_2}y^{\beta_3}\right) \\ &+ 3\tilde{P}_4(\mathbf{x},\mathbf{y})^{-1/2}\tilde{D}_{\alpha\beta\alpha_1\alpha_2}y^{\alpha_1}y^{\alpha_2}. \end{aligned} \tag{7}$$

This means that at each position $\mathbf{x}$ and for each choice of $\mathbf{y}$, we can obtain the corresponding local metric. Figure 2 illustrates the directional dependence of the Finsler metric for a given $\tilde{P}_4(\mathbf{x},\mathbf{y})$ profile. Three different tensors $g_{\alpha\beta}(\mathbf{x},\mathbf{y})$ are obtained for the same profile, i.e., position $\mathbf{x}$, by changing $\mathbf{y}$. The metric tensors are necessary to be able to compute geodesics in the Finsler geometry framework.

3 Geodesic Equations for the Finsler Metric

Analogous to the Riemannian case, in Finsler metric geodesics are the curves that minimize the length between fixed endpoints. We consider a bounded curve $\mathscr{C}$ with parametrization $\mathbf{x} = \boldsymbol{\chi}(t)$, $a \le t \le b$, where t denotes the arc-length. The length of $\mathscr{C}$ is given by,

$$J[\chi] = \int_a^b F(\chi(t), \dot{\chi}(t))\mathrm{d}t \tag{8}$$

where $\dot{\chi}(t) = \frac{\mathrm{d}\chi(t)}{\mathrm{d}t}$. It can be shown that the necessary condition to minimize the length functional (8) is the set of Euler-Lagrange equations [18,23,26],

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial F}{\partial y^\alpha}\right) - \frac{\partial F}{\partial x^\alpha} = 0, \tag{9}$$

where $y^\alpha = \dot{x}^\alpha$. We will derive the geodesic equations from Eq. (9). First, straightforward application of the chain rule gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial F^2}{\partial y^\alpha}\right) - \frac{\partial F^2}{\partial x^\alpha} = \frac{1}{F}\frac{\mathrm{d}F}{\mathrm{d}t}\frac{\partial F^2}{\partial y^\alpha}. \tag{10}$$

Using that $\frac{\mathrm{d}F}{\mathrm{d}t} = 0$ (arclength parametrization), the above equation simplifies to,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial F^2}{\partial y^\alpha}\right) - \frac{\partial F^2}{\partial x^\alpha} = 0. \tag{11}$$

Once more applying the chain rule and substitution of (5) gives

$$2g_{\alpha\beta}\ddot{x}^\beta + \frac{\partial^2 F^2}{\partial y^\alpha \partial x^\beta} y^\beta - \frac{\partial F^2}{\partial x^\alpha} = 0. \tag{12}$$

Multiplying Eq. (12) with the inverse $g^{\gamma\alpha}$ gives

$$\ddot{x}^\alpha + 2G^\alpha(\mathbf{x}, \dot{\mathbf{x}}) = 0, \tag{13}$$

where G^α are the so-called geodesic coefficients defined by

$$G^\alpha(\mathbf{x}, \mathbf{y}) = \frac{1}{4} g^{\alpha\beta}(\mathbf{x}, \mathbf{y}) \left(\frac{\partial^2 F^2(\mathbf{x}, \mathbf{y})}{\partial y^\beta \partial x^\gamma} y^\gamma - \frac{\partial F^2(\mathbf{x}, \mathbf{y})}{\partial x^\beta} \right). \tag{14}$$

It is often useful from a computational point of view to write the geodesic Eq. (13) in an alternative form. To this purpose we introduce the Christoffel symbols (of second kind) $\Gamma^\alpha_{\beta\gamma}$ defined as

$$\Gamma^\alpha_{\beta\gamma}(\mathbf{x}, \mathbf{y}) = \frac{1}{2} g^{\alpha\kappa}(\mathbf{x}, \mathbf{y}) \left(\frac{\partial g_{\beta\kappa}}{\partial x^\gamma} + \frac{\partial g_{\gamma\kappa}}{\partial x^\beta} - \frac{\partial g_{\beta\gamma}}{\partial x^\kappa} \right). \tag{15}$$

Note that compared to the Riemannian case, $\Gamma^\alpha_{\beta\gamma}$ are functions of both space and orientation. To reformulate the geodesic equations, we essentially rewrite G^α

in terms of the formal Christoffel symbols [9]. Indeed, we have the following proposition:

Proposition 1. *The geodesic coefficients G^α defined in (14), are related to the Christoffel symbols $\Gamma^\alpha_{\beta\gamma}$ defined in relation (15) as follows*

$$2G^\alpha = \Gamma^\alpha_{\beta\gamma} y^\beta y^\gamma. \tag{16}$$

Note that this proposition implies the derivatives of F^2 are replaced by derivatives of $g_{\alpha\beta}$. Before we prove relation (16) we need two lemmas. The first one concerns the Cartan tensor $C_{\alpha\beta\gamma}(\mathbf{x}, \mathbf{y})$ defined as:

$$C_{\alpha\beta\gamma}(\mathbf{x}, \mathbf{y}) = \frac{1}{4} \frac{\partial^3 F^2(\mathbf{x}, \mathbf{y})}{\partial y^\alpha \partial y^\beta \partial y^\gamma}.$$

Lemma 1. *The Cartan tensor $C_{\alpha\beta\gamma}(\mathbf{x}, \mathbf{y})$ satisfies*

$$C_{\alpha\beta\gamma}(\mathbf{x}, \mathbf{y}) y^\gamma = 0. \tag{17}$$

Proof. The Riemann-Finsler metric $g_{\alpha\beta}(\mathbf{x}, \mathbf{y})$ inherits the following homogeneity property from the defining property $F(\mathbf{x}, \lambda\mathbf{y}) = \lambda F(\mathbf{x}, \mathbf{y})$ for all $\lambda \geq 0$:

$$g_{\alpha\beta}(\mathbf{x}, \lambda\mathbf{y}) = g_{\alpha\beta}(\mathbf{x}, \mathbf{y}).$$

Differentiation with respect to λ and setting $\lambda = 1$ yields

$$\frac{\partial g_{\alpha\beta}(\mathbf{x}, \mathbf{y})}{\partial y^\gamma} y^\gamma = 0.$$

Consequently, the Cartan tensor satisfies

$$C_{\alpha\beta\gamma}(\mathbf{x}, \mathbf{y}) y^\gamma = \frac{1}{2} \frac{\partial g_{\alpha\beta}(\mathbf{x}, \mathbf{y})}{\partial y^\gamma} y^\gamma = 0.$$

Lemma 2. *The Riemann-Finsler metric tensor satisfies the relations*

$$g_{\alpha\beta} y^\alpha y^\beta = F^2, \tag{18}$$

$$\frac{\partial}{\partial y^\gamma} \left(g_{\alpha\beta} y^\alpha y^\beta \right) = 2 g_{\gamma\alpha} y^\alpha. \tag{19}$$

Proof. From the homogeneity of F, we can derive the following relations

$$\frac{\partial F}{\partial y^\alpha} y^\alpha = F, \quad \frac{\partial^2 F}{\partial y^\alpha \partial y^\beta} y^\beta = 0.$$

First we prove relation (18). Using definition (5) we see that

$$\begin{aligned} g_{\alpha\beta} y^\alpha y^\beta &= \frac{1}{2}\frac{\partial}{\partial y^\alpha}\left(\frac{\partial F^2}{\partial y^\beta}\right) y^\alpha y^\beta \\ &= \frac{\partial}{\partial y^\alpha}\left(F\frac{\partial F}{\partial y^\beta}\right) y^\alpha y^\beta \\ &= \left(\frac{\partial F}{\partial y^\alpha}\frac{\partial F}{\partial y^\beta} + F\frac{\partial^2 F}{\partial y^\alpha \partial y^\beta}\right) y^\alpha y^\beta \\ &= \frac{\partial F}{\partial y^\alpha}\frac{\partial F}{\partial y^\beta} y^\alpha y^\beta = F^2, \end{aligned}$$

where we used the relations above. Next, to show (19), we have,

$$\begin{aligned} \frac{\partial}{\partial y^\gamma}\left(g_{\alpha\beta} y^\alpha y^\beta\right) &= \frac{\partial g_{\alpha\beta}}{\partial y^\gamma} y^\alpha y^\beta + g_{\alpha\beta}\delta_{\alpha\gamma} y^\beta + g_{\alpha\beta} y^\alpha \delta_{\beta\gamma} \\ &= 2C_{\alpha\beta\gamma} y^\alpha y^\beta + 2g_{\gamma\beta} y^\beta \end{aligned}$$

using symmetry of $g_{\alpha\beta}$. Furthermore, using Lemma 1, the first term on the right vanishes. □

Proof of Proposition 1. Given the lemmas above, the following derivation allows us to rewrite G^α in terms of Christoffel symbols, viz.,

$$\begin{aligned} 2G^\alpha &= \frac{1}{2} g^{\alpha\beta}\left(\frac{\partial}{\partial x^\gamma}\frac{\partial F^2}{\partial y^\beta} y^\gamma - \frac{\partial F^2}{\partial x^\beta}\right) \\ &= \frac{1}{2} g^{\alpha\beta}\left(\frac{\partial}{\partial x^\gamma}\frac{\partial}{\partial y^\beta}\left(g_{\lambda\mu} y^\lambda y^\mu\right) y^\gamma - \frac{\partial g_{\lambda\mu}}{\partial x^\beta} y^\lambda y^\mu\right), \end{aligned}$$

where we substituted expression (18) for F^2. Next combining this relation with (19), we obtain

$$\begin{aligned} 2G^\alpha(\mathbf{x},\mathbf{y}) &= \frac{1}{2} g^{\alpha\beta}\left(\frac{\partial}{\partial x^\gamma}\left(2g_{\beta\lambda} y^\lambda\right) y^\gamma - \frac{\partial g_{\lambda\mu}}{\partial x^\beta} y^\lambda y^\mu\right) \\ &= \frac{1}{2} g^{\alpha\beta}\left(\frac{\partial g_{\lambda\beta}}{\partial x^\mu} + \frac{\partial g_{\mu\beta}}{\partial x^\lambda} - \frac{\partial g_{\lambda\mu}}{\partial x^\beta}\right) y^\lambda y^\mu \\ &= \Gamma^\alpha_{\lambda\mu} y^\lambda y^\mu, \end{aligned}$$

which completes the derivation of (16). □

Finally, substituting (16) in (13) gives

$$\ddot{x}^\alpha + \Gamma^\alpha_{\beta\gamma}\dot{x}^\beta \dot{x}^\gamma = 0, \tag{20}$$

which is the desired alternative form of the geodesic equations.

4 Numerical Model

In order to develop a numerical model, we will focus on a specific HARDI model. We have chosen the fourth-order tensor representation of the Orientation Distribution Function (ODF) [1, 11, 12, 28]. From the DWI data, we compute the ODF as was introduced by Tristán-Vega et al. [28]. In the original literature, the ODF is represented by spherical harmonics. A change of basis allows us to transform the spherical harmonics representation of the ODF to higher order tensor representation. To compute the Finsler metric, we need to compute the spherical inversion of the ODF at each voxel. To achieve this, we apply relation (3) for $m > 15$ different orientations $\mathbf{y}$ on the unit sphere, giving rise to an over-determined system $Y\tilde{\mathbf{d}} = \mathbf{b}$, where $\tilde{\mathbf{d}}$ contains the coefficients; see Table 1. To evenly sample the orientations $\mathbf{y}$ on the sphere, we apply the icosahedron tessellation on the unit sphere. In this chapter, we restrict ourselves to the normalized ODF, hence $\overline{P}_4(\mathbf{x}, \mathbf{y}) = 1$. To compute the solution of this system, least-squares fitting is applied. We rewrite (3) in the form of normal equation $Y^{\mathrm{T}}Y\tilde{\mathbf{d}} = Y^{\mathrm{T}}\mathbf{b}$ where

$$Y = \begin{pmatrix} y_1^1 y_1^1 y_1^1 y_1^1 & \cdots & y_1^3 y_1^3 y_1^3 y_1^3 \\ y_2^1 y_2^1 y_2^1 y_2^1 & \cdots & \cdots \\ \vdots & \vdots & \vdots \\ y_m^1 y_m^1 y_m^1 y_m^1 & \cdots & y_m^3 y_m^3 y_m^3 y_m^3 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} P_4(\mathbf{x}, \mathbf{y}_1)^{-1} \\ \vdots \\ P_4(\mathbf{x}, \mathbf{y}_m)^{-1} \end{pmatrix}. \tag{21}$$

Note that the lower indices of $\mathbf{y}$ indicate the choice of direction and m is the number of gradient directions to sample the ODF profile; in our computations we use $m = 72$. The solution $\tilde{\mathbf{d}}$ is computed using Cholesky factorization. This gives 15 coefficients for the inverted ODF profile. We can fit the new profile over the sphere using the new coefficients $\tilde{\mathbf{d}}$; see Figs. 3 and 4.

Let us introduce $u^\gamma(t) := \dot{x^\gamma}(t)$ for $\gamma = 1, 2, 3$, then we can rewrite system (20) as follows

$$\begin{aligned} \dot{x}^\alpha &= u^\alpha, \\ \dot{u}^\alpha &= -\Gamma^\alpha_{\beta\gamma} u^\beta u^\gamma, \end{aligned} \tag{22}$$

with the $\Gamma^\alpha_{\beta\gamma}$ defined in (15). To solve the system of equations (22) we follow a similar algorithm proposed by Sepasian et al. [24] for the Riemannian metric. Consider $\big(x^1(0), x^2(0), x^3(0)\big)$ a point as the given initial location in the domain and $\big(u^1(0), u^2(0), u^3(0)\big)$ as the initial direction. We compute the solution to (22) for the given initial position and multiple directions using sophisticated ODE solvers such as the fourth order explicit Runge-Kutta method. This gives us a set of geodesics connecting the given initial point to a set of points on the boundary.

The computational domain is discretized uniformly with grid size h and grid points $\mathbf{x}_{ijk} = (x_i^1, x_j^2, x_k^3) = h(i, j, k)$ for $i = 0, 2, 3, \ldots, N-1$, where N is number of grid points in each spatial direction. For simplicity we take the

Table 1 Ordering of higher order tensor coefficients

Tensor element	Coeff. of HOT	Tensor element	Coeff. of HOT
1	D_{1111}	8	D_{1223}
2	D_{1112}	9	D_{1233}
3	D_{1113}	10	D_{1333}
4	D_{1122}	11	D_{2222}
5	D_{1123}	12	D_{2223}
6	D_{1133}	13	D_{2233}
7	D_{1222}	14	D_{2333}
		15	D_{3333}

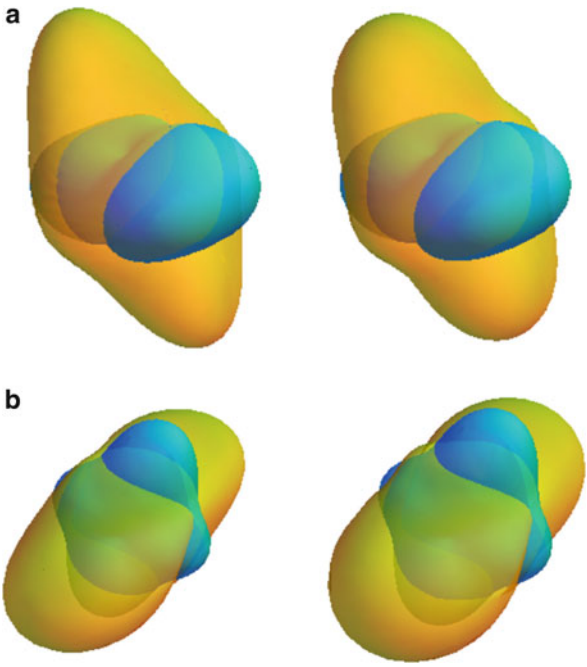


Fig. 3 ODF (*Blue*) and its inverse (*Yellow*) using least-squares fit (*Left*) and analytic inversion (*Right*). (**a**) ODF and its inverse for a single fiber profile. (**b**) ODF and its inverse for a crossing fiber profile

number of grid points equal in all directions. For each grid point, we assign the 15 coefficients of inverted HOT $\tilde{\mathbf{D}}$ (see Fig. 5 top).

We approximate the derivatives of $g_{\alpha\beta}(\mathbf{x}, \mathbf{y})$ at each grid point by the standard second order central difference scheme, for example,

$$\frac{\partial g_{\alpha\beta}}{\partial x^1}(x_i^1, x_j^2, x_k^3, \mathbf{y}) \approx \frac{1}{2h}\left(g_{\alpha\beta}(x_{i+1}^1, x_j^2, x_k^3, \mathbf{y}) - g_{\alpha\beta}(x_{i-1}^1, x_j^2, x_k^3, \mathbf{y})\right). \quad (23)$$

Second order one-sided differences are applied when the grid points are situated on the boundary, e.g,

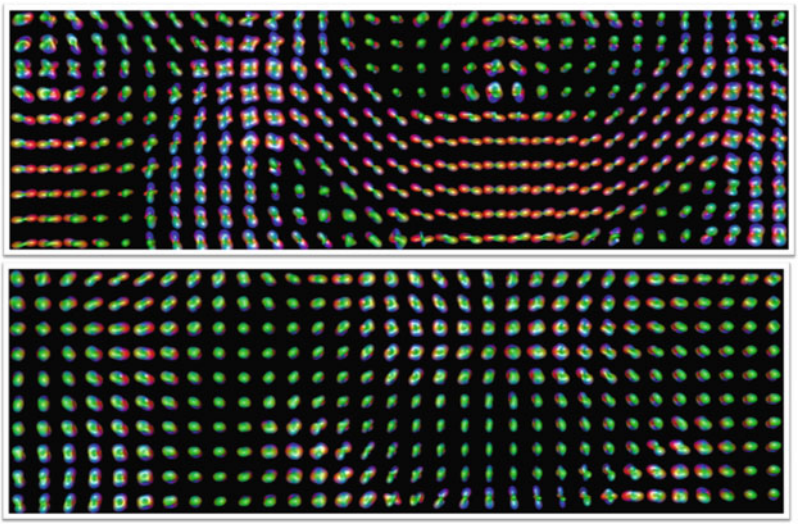


Fig. 4 Fourth order ODF field (*top image*) and its inverted field (*bottom image*) from part of the centrum semiovale

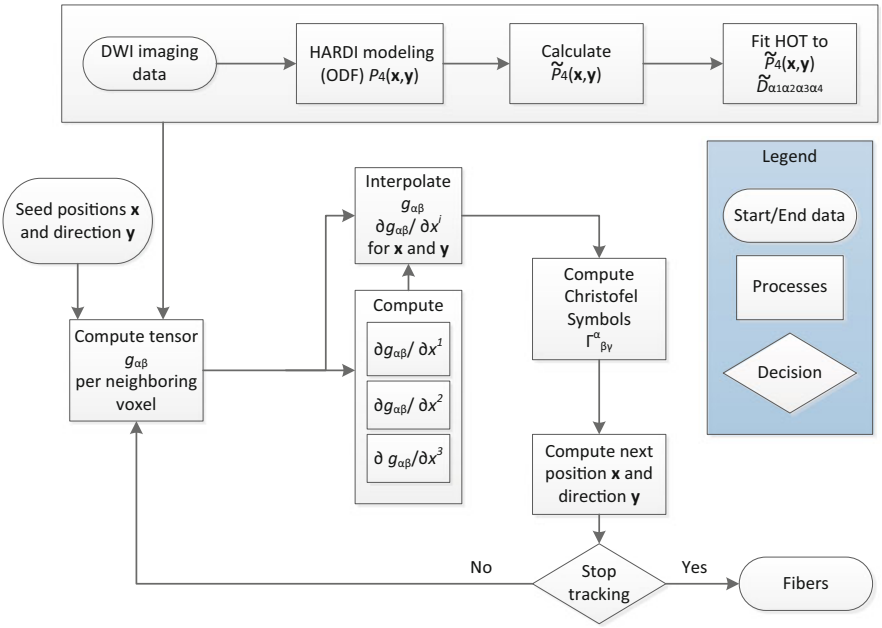


Fig. 5 Flowchart for the ray tracing algorithm to reconstruct geodesics in a Finsler space

$$\frac{\partial g_{\alpha\beta}}{\partial x^1}(x_0^1, x_j^2, x_k^3, \mathbf{y}) \approx$$
$$\frac{1}{2h}\left(-3g_{\alpha\beta}(x_0^1, x_j^2, x_k^3, \mathbf{y}) + 4g_{\alpha\beta}(x_1^1, x_j^2, x_k^3, \mathbf{y}) - g_{\alpha\beta}(x_2^1, x_j^2, x_k^3, \mathbf{y})\right). \quad (24)$$

Note that similar expressions hold for derivatives with respect to x^2 and x^3. Note the dependence of these relations on the argument $\mathbf{y}$. Solving the ODE system gives the solution at points that are not necessarily located in the grid points. Therefore, the value of the metric and its derivatives are not defined there, and we apply trilinear interpolation at any point in the domain where the value of the metric is not available. Initial vectors are uniformly distributed over the unit sphere using the vertices of regular symmetrical polyhedra. The integration of geodesics continues till they hit the boundary of the computational domain. Once all geodesics are computed for the initial seed points and given initial directions, one can select the fibers by selecting the regions of interests and filter all geodesics that pass through both selected regions. Geodesics are computed until they meet one of the boundaries, therefore to determine the fiber connecting two given regions we apply the line-plane intersection. This allows us to cut off the geodesics once they enter one of the selected regions. The flow chart in Fig. 5 summarizes the ray-tracing tractography algorithm for HARDI. The algorithm has the following stages:

1. Compute the coefficients of the inverted ODFs using the least-squares fit as described in this section (see Fig. 5 top).
2. Fetch an initial seed point and direction, which becomes the current position and direction.
3. Find the eight neighboring voxels forming the cell of the current position and compute for each neighboring voxel the metric tensor according to Eq. (7).
4. Compute the derivatives at the neighboring voxels.
5. Compute the metric tensor and its derivatives at the current position using trilinear interpolation.
6. Compute the Christoffel symbols, as defined in Eq. (15).
7. Compute the next position and direction of the fiber, from (22)
8. Repeat step 3–6 until the stopping criterion (i.e., meeting the domain boundaries) is fulfilled.

5 Results

In this section we present the first results we have obtained with the ray-tracing tractography algorithm described in this chapter. We generated synthetic tensor fields simulating two fiber bundles crossing at angles 90° and 60°. To generate the tensors in the crossing area the Gaussian mixture model introduced by Tuch et al. [29] has been used. The signal is obtained simulating a b-value of 1,000 s/mm^2 and voxel resolution $1 \times 1 \times 1$ mm. Riccian noise with $SNR = 15$ is added to the images.

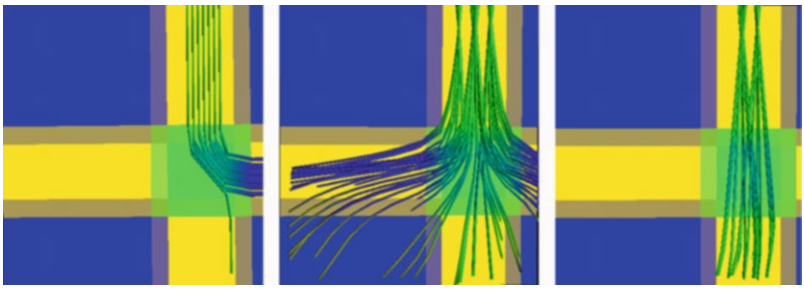


Fig. 6 90° crossing. From left to right: streamline tractography for DTI, geodesic ray tracing for DTI, geodesic ray tracing for HARDI

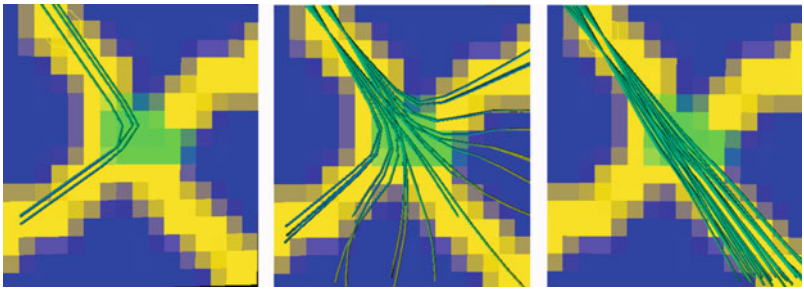


Fig. 7 Rotated 90° crossing. From left to right: streamline tractography for DTI, geodesic ray tracing for DTI, HARDI multi-valued geodesic

Results for the 90° crossing angle are shown in Fig. 6. The behavior of the streamline (left), DTI ray-tracing [25] (middle) and HARDI ray-tracing (right) tractography are illustrated. The results illustrate that streamlines have a preference for curving instead of going straight. The HARDI ray-tracing tractography gives less spreading of the reconstructed fiber bundles compared to the DTI ray-tracing method. Figure 7 shows the behavior of these algorithms with respect to rotation. The crossing angle here remains 90° and the data is rotated 55°. Figure 8 illustrate fiber bundles for the 60° crossing angle. We see that for a sharper angle the algorithm is still capable of capturing crossing fibers.

In order to provide preliminary results of HARDI ray-tracing tractography for real data, Fig. 9 shows the tractography result using a $10 \times 10 \times 30$ cube cut from the human data set obtained by a Philips scanner with resolution $2 \times 2 \times 2$ mm, b-value of 1,000 s/mm^2 and 128 gradient directions. The fiber tracking stops in both cases if the fibers meet one of the boundaries. These preliminary results show the capability of the tractography algorithm for capturing realistic fiber tracts in the area of corona radiata (CR), corpus callosum (CC) and cingulum (CG). However, these are preliminary results and future research and experiments are needed to fully explore the capabilities of the presented method.

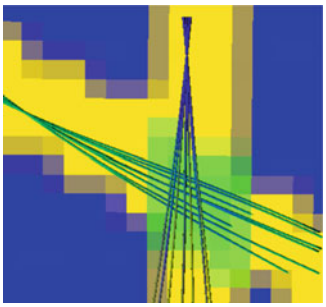


Fig. 8 Geodesic ray tracing for HARDI using 60° crossing

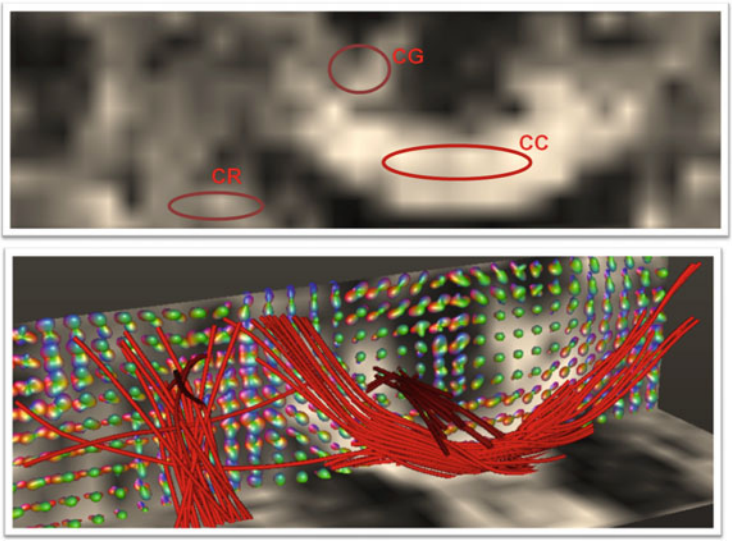


Fig. 9 Tractography results in the crossing region of Fig. 4. The *top image* indicates the seed regions in a fractional anisotropy map, the *bottom image* shows the ray-tracing tractography results for CC, CR and CG

6 Conclusion

In this chapter, we presented a new tractography algorithm for HARDI data. Our method is based on computing geodesics in the Finsler metric as an extension of Riemannian metric. The Finsler metric is defined as function of position and direction using HARDI data at each voxel. We made a heuristic choice concerning the mapping of the HARDI data into a Finsler metric. For future work, it would be interesting to study other possibilities. Compared to other existing geodesic based HARDI tractography methods, we computed multi-valued solutions of geodesic equations instead of single valued viscosity solutions. We presented the theory and

showed the potential of our method for capturing crossing fibers. Results of synthetic data and very preliminary human brain data results showed the applicability of the method. These are encouraging results that need further experimentation to explore its possibilities and compare it to other existing fiber tracking methods.

References

1. Aganj, I., Lenglet, C., Sapiro, G.: ODF reconstruction in q-ball imaging with solid angle consideration. In: The 6th IEEE International Symposium on Biomedical Imaging (ISBI), Boston, pp. 1398–1401 (2009)
2. Alexander, D.C., Barker, G., Arridge, S.: Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. Magn. Reson. Med. **48**(2), 331–340 (2002)
3. Astola, L., Florack, L.: Finsler geometry on higher order tensor fields and applications to high angular resolution diffusion imaging. Int. J. Comput. Vis. **92**, 325–336 (2011)
4. Basser, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J. Magn. Reson. **111**(1), 209–219 (1996)
5. Behrens, T.E.J., Berg, H.J., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? NeuroImage **34**(1), 144–155 (2007)
6. Blair, D.E.: Inversion Theory and Conformal Mapping. American Mathematical Society, Providence (2000)
7. Lenglet, C., Deriche, R., Faugeras, O.D.: Inferring white matter geometry from diffusion tensor MRI: application to connectivity mapping. In: Proceedings of 8th European Conference on Computer Vision, Prague, pp. 127–140 (2004)
8. Campbell, J.S.: Diffusion imaging of white matter fibre tracts. PhD thesis, McGill University, Montreal (2004)
9. Chern, S., Shen, Z.: Lectures on Finsler geometry. Nankai Tracts Math. **6** (2003)
10. Morris, D.M., Embleton, K.V., Parker, G.J.: Probabilistic fibre tracking: differentiation of connections from chance events. NeuroImage **42**(4), 1329–1339 (2008)
11. Descoteaux, M.: High angular resolution diffusion MRI: from local estimation to segmentation and tractography. Ph.D. Thesis, Universite de Nice, Sophia Antipolis (2008)
12. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast and robust analytical q-ball imaging. Magn. Reson. Med. **58**(3) (2007)
13. Descoteaux, M., Deriche, R., Knoesche, T., Anwander, A.: Deterministic and probabilistic tractography based on complex fiber orientation distributions. IEEE Trans. Med. Imaging **2**(28), 269–286 (2008)
14. Donnell, L.O., Haker, S., Westin, C.F.: New approaches to estimation of white matter connectivity in diffusion tensor MRI: Elliptic PDEs and geodesics in a tensor-warped space. In: International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'02, Tokyo, vol. 2488, pp. 459–466 (2002)
15. Hagmann, P., Reese, T., Tseng, W., Meuli, R., Thiran, J., Wedeen, V.: Diffusion spectrum imaging tractography in complex cerebral white matter: an investigation of the centrum semiovale. In: ISMRM, Kyoto, vol. 12. International Society for Magnetic Resonance in Medicine (2004)
16. Lenglet, C., Prados, E., Pons, J.P., Deriche, R., Faugeras, O.: Brain connectivity mapping using Riemannian geometry, control theory and PDEs. SIAM J. Imaging Sci. (SIIMS) **2**(2), 285–322 (2009)
17. Melonakos, J., Pichon, E., Angenent, S., Tannenbaum, A.: Finsler active contours. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 412–423 (2008)

18. Mo, X.: An Introduction to Finsler Geometry. Volume 1 of Peking University Series in Mathematics. World Scientific, Singapore (2006)
19. Ozarslan, E., Shepherd, T., Vemuri, B., Blackband, S., Mareci, T.: A nonparametric reconstruction and its matrix implementation for the diffusion orientation transform (dot). In: The 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro (ISBI), Arlington, pp. 85–88 (2006)
20. Parker, G.J.M., Alexander, D.C.: Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. Philos. Trans. R. Soc. Lond. B: Biol. Sci. **360**(1457), 893–902 (2005)
21. Péchaud, M., Descoteaux, M., Keriven, R.: Brain connectivity using geodesics in HARDI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'09, London, pp. 482–489 (2009)
22. Pichon, E., Westin, C., Tannenbaum, A.: A Hamilton-Jacobi-Bellman approach to high angular resolution diffusion tractography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI'05, Palm Springs. Lecture Notes in Computer Science, pp. 180–187 (2005)
23. Rund, H.: The Hamilton-Jacobi Theory in the Calculus of Variations. Robert E. Krieger Publishing, Huntington (1973)
24. Sepasian, N., ten Thije Boonkkamp, J., ter Haar Romeny, B., Vilanova, A.: Multivalued geodesic ray-tracing for computing brain connections using diffusion tensor imaging. SIAM J. Imaging Sci. **5**(2), 483–504 (2012)
25. Sepasian, N., ten Thije Boonkkamp, J., Vilanova, A., ter Haar Romeny, B.: Multi-valued geodesic based fiber tracking for diffusion tensor imaging. In: MICCAI'09, Diffusion Modeling and the Fiber Cup Workshop, London, vol. 1, pp. 6–13 (2009)
26. Shen, Z.: Lectures on Finsler Geometry. World Scientific, Singapore (2001)
27. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. NeuroImage **35**(4), 1459–1472 (2007)
28. Tristán-Vega, A., Westin, C.F., Aja-Fernández, S.: Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging. NeuroImage **47**(2), 638–650 (2009)
29. Tuch, D., Reese, T., Wiegell, M., Makris, N., Belliveau, J.W., Wedeen, V.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magn. Reson. Med. **48**, 577–582 (2002)
30. Tuch, D.S.: Q-ball imaging. Magn. Reson. Med. **52**(6), 1358–1372 (2004)
31. Wedeen, V.J., Hagmann, P., Tseng, W.Y.I., Reese, T.G., Weisskoff, R.M.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn. Reson. Med. **54**(6), 1377–1386 (2005)

Part IV
Tensor Signal Processing

Kernel-Based Morphometry of Diffusion Tensor Images

Madhura Ingalhalikar, Parmeshwar Khurd, and Ragini Verma

Abstract Voxel-based group-wise statistical analysis of diffusion tensor imaging (DTI) is an integral component in most population-based neuroimaging studies such as those studying brain development during infancy or aging, or those investigating structural differences between healthy and diseased populations. The majority of studies using DTI limit themselves by testing only certain properties of the tensor that mainly include anisotropy and diffusivity. However, the pathology under study may affect other aspects like the orientation information provided by the tensors. Therefore, for detecting subtle pathological changes it is important to perform group-wise testing on the whole tensor, which encompasses the changes in anisotropy, diffusivity and orientation. This is rendered challenging by the fact that conventional linear statistics cannot be applied to tensors. Moreover, the pathology over the population is unknown and could be non-linear, further complicating the group-based statistical analysis. This chapter gives a perspective on performing voxel-wise morphometry of tensor data using kernel-based approach. The method is referred as Kernel-based morphometry (KBM) as it models the tensor distribution using kernel principal component analysis (kPCA), which linearizes the data in high dimensional space. Subsequently a Hotelling T^2 test is performed on the high dimensional kernelized data to determine statistical group differences. We apply this method on simulated and real datasets and show that KBM can effectively identify the underlying tensorial distribution. Thus it can potentially elucidate pathology-induced population differences, thereby establishing a kernelized full tensor framework for population studies.

M. Ingalhalikar (✉) • P. Khurd • R. Verma
Department of Radiology, Section of Biomedical Image Analysis, University of Pennsylvania, Philadelphia, PA 19104, USA
e-mail: Madhura.Ingalhalikar@uphs.upenn.edu; Ragini.Verma@uphs.upenn.edu

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_10,

1 Introduction

Diffusion tensor imaging (DTI) is an important MRI modality for studying white matter connectivity and organization non-invasively [6, 32]. There exists a wide variety of applications of DTI that include investigating neurological and psychiatric disorders [2,21,25], structure-function relationships [11,20,38], evaluation of brain connections and creating the structural brain connectome [8, 19], and as a tool to assist in computer guided surgery and treatment planning [13,29].

Clinical investigation of pathology-induced changes requires a group-based statistical analysis of DT images which can identify regional differences between controls and patients. Voxel-based morphometry (VBM) [4] is a popular form of statistical analysis and has been widely adopted by the neuroimaging community. VBM has an advantage over region-of-interest (ROI) based analysis as it does not require an a priori hypothesis regarding regions affected by disease. Conventionally, VBM on DTI images is performed via analyzing the scalar maps of anisotropy and/or diffusivity that are computed from the tensor and then spatially normalized to a common template. A statistical voxel-wise p-value map is then computed from these scalar maps with the application of standard tests (e.g. t-tests) for statistical inference. Commonly used scalar images for such analyses include the fractional anisotropy (FA) and the apparent diffusion coefficient (ADC) [28,30,37]. The main disadvantage of these methods is that they do not use the complete information available in the DTI dataset, but rather make an a priori assumption that group differences will affect only a particular aspect of the tensors which is usually quantified through scalar indices like FA. Moreover, the combined results from different scalar maps may be difficult to interpret as they can potentially contain spatially overlapping patterns. Tensors provide both shape information (in the form of eigenvalues) that are captured in scalar maps of anisotropy and diffusivity measures computed from the tensor data and underlying fiber orientation information in the form of eigenvectors. However, voxel-wise statistical analysis of the tensors is complicated as the tensor is a 3×3 positive definite symmetric matrix located at each voxel and has an underlying non-linear manifold structure [3, 14, 31]. Furthermore, disease-induced changes between the subjects may not be linear adding to the complexity of the group-based statistical analysis.

Non-scalar features [34,35,42] such as the principal eigen-directions (PD) of the tensors have also been used in some analyses. In [34] a Bipolar-Watson model was introduced for analysis of PD's. This model takes into account the symmetric nature of the tensor while performing the statistical analysis and assumes the underlying distribution to be a Wishart distribution. However, in the regions where white matter fibers cross, the tensors are oblate in nature and therefore applying statistics to PD's may demonstrate ambiguous results.

A few methods have attempted on analyzing tensors in the VBM type setting. For example, work by Li et al. performed tensor regression analysis [27] where a linear regression model was assumed that imposed distributional assumptions on the tensors under consideration. In some other works, [15, 26] tensors were

characterized using Riemannian symmetric spaces. In [3] a simple and efficient Riemannian framework based on Log-Euclidean (LE) transform was introduced. Such methods rely upon the assumption that the tensors around a given voxel from various subjects belong to a principal geodesic sub-manifold and that these tensors obey a normal distribution on that sub-manifold. The basic principle of these methods is sound, namely that statistical analysis of tensors must be restricted to the appropriate manifold of positive definite symmetric tensors, which is known to be a cone embedded in $\Re^6$. However, there is no guarantee that the representations of the tensors on this sub-manifold will have normal distributions since the pathology imposes its own structure and the tensors measured at a given voxel, from n subjects, typically lie on a much more restricted submanifold of the space of symmetric positive definite matrices. A novel approach was suggested by Verma et al., where a manifold learning method (Isomap) was employed for tensor analysis [40]. The focus of this work was on learning embeddings (or features) parameterizing the underlying manifold structure of the tensors. The learned features belonged to a low-dimensional linear manifold parameterizing the higher-dimensional tensor manifold and were subsequently used for group-wise statistical analysis. In general, manifold learning approaches [10] may be used to estimate the embedding of the manifold that represents the tensor measurements fairly well, however, depending on the number of samples used to learn the underlying manifold structure, it may not always be possible to determine the structure or validate its correctness or they may fail to estimate the probability distribution (non-Gaussian) on the (flattened) manifold itself.

This chapter is aimed towards providing a paradigm for voxel-wise tensor statistics by determining the underlying statistical distribution of the data and using this distribution for subsequent voxel-wise analysis. To achieve this, a novel method named kernel-based morphometry (KBM) is developed which demonstrating that it can accurately estimate the underlying distribution of the tensor data compared to other existing methods. A standard statistical test is then performed on these projections with appropriate testing for multiple comparisons.

2 Kernel Based Approach to Group-Wise Voxel Based DTI Statistical Analysis

Group-wise voxel-based statistical analysis of DTI data involves spatially normalizing all DT images to a common DTI template using a suitable technique [12, 22, 43, 46] and then using an appropriate voxel-wise statistical test to infer regional differences between groups based upon the tensors at (or around) a voxel. In this kernel based approach, the underlying distribution of the data at each voxel is determined and subsequently statistics are performed.

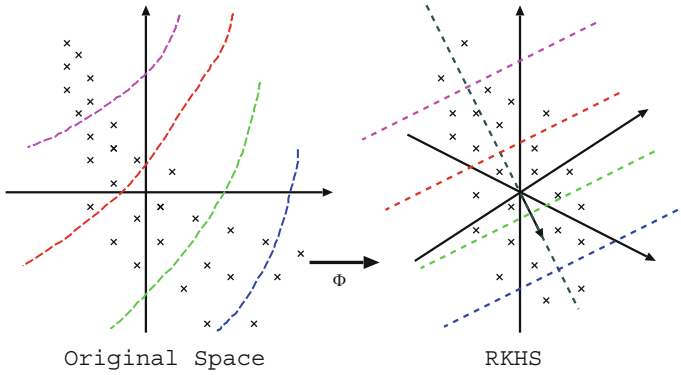


Fig. 1 Kernel-based projections: The mapping $\boldsymbol{\phi}$ takes points (marked with *crosses*) from the original non-linear space to the linearized RKHS. Hyperplanes having constant projections onto a vector in the RKHS become curved lines in the original space. Such curved lines can give us important insight into how the corresponding RKHS projection parameterizes the original points

Kernel Principal Component Analysis (kPCA) [33] is a suitable method for learning the underlying data distribution. The common idea in any of the kernel-based techniques is to transform the samples into a higher-dimensional reproducible kernel Hilbert space (RKHS). Samples can be expressed using an appropriate kernel in a higher dimensional space using the well known "kernel trick" [33]. The non-linear hypersurfaces in the original space are mapped into hyperplanes in RKHS. These hyperplanes separate the given samples linearly in RKHS which is equivalent to a non-linear separation in original space. Subsequently, statistical operations can be performed in this "kernelized" space. Thus, in case of DTI, at each voxel, the intensities are kernelized to hyperplanes in the RKHS. Figure 1 illustrates the idea behind obtaining such components. Since these components are linear in the RKHS, linear tests for statistical inference, such as the Hotelling T^2 test, can be reliably applied to these projections in order to identify separation between groups.

Before presenting the kPCA technique in detail, here is a brief note on our mathematical conventions: Vectors are denoted by bold-faced lower case letters, e.g. $\mathbf{x}$, and matrices by upper-case letters, e.g. S for tensor matrix with U as the eigenvector matrix and D as the diagonal matrix containing the eigenvalues of the tensor. Vector of all 1's is denoted by $\mathbf{e}$ while the identity matrix is denoted by I. Vector of all 1's in m-dimensional space is represented by $\mathbf{e}_m$. Matrix transpose and the matrix inverse are denoted by superscripts ^{T} and $^{-1}$ respectively. The sample mean of a set of vectors $\{\mathbf{x}_i, i = 1, \cdots, K\}$ is represented as $\bar{\mathbf{x}}$, while the inner product of two vectors $\mathbf{x}_i, \mathbf{x}_j$ is denoted by $< \mathbf{x}_i, \mathbf{x}_j >$. Group-wise study includes a statistical analysis of the DT images of N subjects, with N_+ subjects in one class (the positive class) and the remaining N_- subjects in a second class (the negative class).

3 Kernel Principal Component Analysis (kPCA)

We now describe the kPCA technique [33] by which one can find a rich linear representation of our voxel-based samples as well as provide an accurate estimate of the probability density underlying these samples. In conventional PCA, principal directions in the vector space of the samples that maximize the variance of the components of the samples along those directions and which also minimize the least-squares representation error for the samples are determined. In kPCA, similar principal eigen-directions in higher-dimensional RKHS are found where can be safely assumed to be normally distributed.

A DT image consists of a 3×3 positive-definite symmetric matrix or tensor S at each voxel in the image. Earlier work presented by Khurd et al. [24] used the diffusion tensor directly as a $6D$ vector that could potentially lead to inaccurate results since the tensors lie on the geodesic sub-manifold which is known to be a cone embedded in $\Re^6$. Therefore, the log-Euclidean form of tensor was employed that retained the key attributes of affine-invariant Riemannian metric, and allowed standard Euclidean computations in the space of matrix logarithms, as was described by Arsigny et al. [3]. A tensor S can be represented as $S = UDU^T$ where matrix U is a matrix of its eigenvectors and D is a diagonal matrix that consists of the three eigenvalues. The log-Euclidean form of the tensor is given by Eq. 1.

$$S^{le} = log(S) = Ulog(D)U^T \tag{1}$$

$$\mathbf{x} = (S^{le}_{xx}, S^{le}_{yy}, S^{le}_{zz}, \sqrt{2}S^{le}_{xy}, \sqrt{2}S^{le}_{xz}, \sqrt{2}S^{le}_{yz})^T \tag{2}$$

A similarity invariant log-Euclidean form of the tensor is computed using Eq. 1 [3]. The 6-dimensional vectors are then obtained using Eq. 2 for all our N subjects $\mathbf{x}_1, \cdots, \mathbf{x}_N$. Let us denote the nonlinear mapping of this vector $\mathbf{x}$ into the Hilbert space by $\boldsymbol{\phi}(\mathbf{x})$, and let us denote the underlying kernel by $k(.,.)$, where $< \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) >= k(\mathbf{x}_i, \mathbf{x}_j)$. Let $\bar{\boldsymbol{\phi}}$ denote the mean of $\boldsymbol{\phi}(\mathbf{x}_1), \cdots, \boldsymbol{\phi}(\mathbf{x}_N)$. Since a principal eigenvector $\mathbf{v}$ in the higher-dimensional Hilbert space lies in the span of the vectors $\boldsymbol{\phi}(\mathbf{x}_i) - \bar{\boldsymbol{\phi}}, i = 1, \cdots, N$, it can be conveniently represented as $\mathbf{v} = \sum_i \alpha_i(\boldsymbol{\phi}(\mathbf{x}_i) - \bar{\boldsymbol{\phi}})$, where $\boldsymbol{\alpha}$ is an N-dimensional vector. Components of any sample along the eigenvector $\mathbf{v}$ can now be conveniently computed using this new representation in the kernel basis.

The entire kPCA procedure is summarized below [10]:

Another alternative to using LE form of tensors and then computing the kernel $K(x_i, x_j)$ is employing directly a kernel $K(log(s_i), log(s_j))$ where s is the original tensor (in a 6 dimensional vector form).

In addition to finding the orthogonal directions of maximal variance in the higher-dimensional RKHS, kPCA also provides an estimate of the probability density underlying the samples. It has been pointed out by Girolami et al. [18] that kPCA with a Gaussian radial basis function kernel amounts to orthogonal series

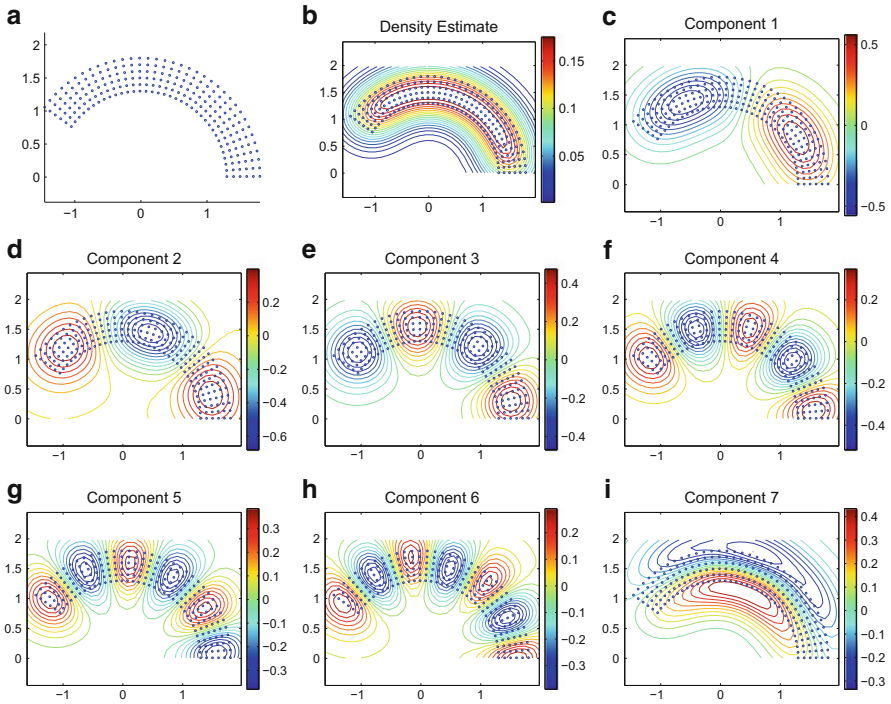


Fig. 2 (**a**) Synthetic dataset, (**b**) Contour plot for kernel probability density estimate, (**c**) Contour plot for 1st kPCA component, (**d**) Contour plot for 2nd kPCA component, (**e**) Contour plot for 3rd kPCA component, (**f**) Contour plot for 4th kPCA component, (**g**) Contour plot for 5th kPCA component, (**h**) Contour plot for 6th kPCA component, and (**i**) Contour plot for 7th kPCA component (please see text for explanation in Sect. 3)

density estimation using Hermite polynomials. Gaussian kernels are frequently employed in alternative kernel-based classifiers such as support vector machines [33]. The advantages of using Gaussian kernel are multifold; it non-linearly maps the samples into RKHS, involves less number of parameters than a polynomial kernel and is known to be robust. The Gaussian σ value was chosen to be based on the average distance between nearest neighbors (NN) x_i and x_j for e.g. $\|x_i - x_j\|$ and our choice was motivated by the desire to obtain meaningful representations for the different kPCA components.

In Sect. 5.1, simulated example is presented (see Fig. 2) where kPCA provides an accurate parametrization of the underlying density of the dataset. We note that the kPCA components constitute a linear representation of the tensors in the RKHS, which considerably simplifies further statistical analysis that will need to be performed on the dataset. An important issue is selecting the number of kPCA

Algorithm 1 kPCA

1. Form the kernel matrix K, where $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j), i = 1, \cdots, N, j = 1, \cdots, N$.
2. Center the kernel matrix to obtain $K_c = (I - \frac{1}{N}\mathbf{e}\mathbf{e}^T)K(I - \frac{1}{N}\mathbf{e}\mathbf{e}^T)$.
3. Eigen-decompose K_c to obtain its eigenvectors $\boldsymbol{\alpha}^{(i)}$ and eigenvalues λ_i, $i = 1, \cdots, N(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N)$.
4. Normalize the eigenvectors $\boldsymbol{\alpha}^{(i)}$ to have length $\frac{1}{\sqrt{\lambda_i}}$ so that the eigenvectors $\mathbf{v}^{(i)}$ in the RKHS have unit length.
5. The i^{th} kPCA component for training sample $\mathbf{x}_k$ is given by:

$$< \boldsymbol{\phi}(\mathbf{x}_k) - \bar{\boldsymbol{\phi}}, \mathbf{v}^{(i)} >= \lambda_i \alpha_k^{(i)}$$

6. For a general test point $\mathbf{x}$, the i^{th} kPCA component is:

$$< \boldsymbol{\phi}(\mathbf{x}) - \bar{\boldsymbol{\phi}}, \mathbf{v}^{(i)} >= \sum_m \alpha_m^{(i)} k(\mathbf{x}, \mathbf{x}_m) - \frac{1}{N} \sum_{m,n} \alpha_m^{(i)} k(\mathbf{x}, \mathbf{x}_n)$$

components used for subsequent statistical analysis. This number can be chosen by looking at the kPCA eigenvalue spectrum and selecting only those eigenvectors that correspond to large eigenvalues. The notion of "large" eigenvalues is empirically defined using a application-specific threshold in one of two ways. The threshold may either specify the minimum energy $\frac{\sum_{i=1}^{L} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$ that should be present in the retained L eigenvalues, or it may specify a minimum value for the ratio of the smallest retained eigenvalue λ_L to the largest retained eigenvalue λ_1. For good discriminatory performance between the groups, the number of kernel PCA components chosen should not exceed the number of samples in either class. Statistical p-value maps are then computed using Hotelling's T^2 statistic on the retained kPCA projections. This procedure is repeated at each voxel to obtain the kernelized version of the tensors at that voxel. These are now vectors in a high-dimensional linear space. Thus linear statistical tools for high dimensional data such as Hotelling's T^2 can then be applied to the retained kPCA components. The resulting p-value map can then be thresholded to obtain regions of interest.

To overcome the multiple comparisons problem associated with voxel-wise analysis, False Discovery Rate (FDR) is implemented. This method controls the expected proportion of falsely rejected hypotheses [9]. The FDR threshold is determined from the observed p-value distribution, and therefore is adaptive to the amount of information in a given dataset [17].

The entire computational procedure for statistical analysis of tensors using kPCA also referred as KBM is summarized below:

This method can also be applied directly to eigenvectors of the tensors for studying the groupwise orientation changes. Performing KBM on a DTI population encompasses changes in scalar maps like FA as well as the orientation.

Algorithm 2 KBM of DTI data:

- **Input**: DTI datasets spatially normalized to a standard template. (N_+ subjects in one group and N_- subjects in the other group).
- **Output**: p-value maps indicating regional differences between the two groups (in general patients and controls).
- *Parametric p-value map:*
  - For each voxel $v = 1, \cdots, V$
    - Compute the log-Euclidean form of tensors from N subjects.
    - Apply kPCA (refer to algorithm 1 above).
    - Select the number of kPCA components (using the energy threshold criterion).
    - Compute Hotelling's T^2 statistic $T^2(v)$ on the kPCA components and the parametric p-value $p(v)$.

  Regions of significance can be identified by controlling the FDR using a suitable p-value threshold. Genovese et al. [17] have recommended the usage of threshold p-values of < 0.1.

4 kPCA Based kFDA

It has been showed that kernel Fisher Discriminant Analysis (kFDA) could be an alternative tensor analysis method to kPCA [24]. kFDA focuses on finding non-linear projections of the tensorial data which can optimally discriminate between two groups. It computes a direction in higher order RKHS such that the projection along this direction maximizes a separability measure known as Rayleigh coefficient (or Fisher discriminant ratio). To quantify the group difference a T^2 test was performed on the kFDA components [24].

It is important to note that kFDA solution uses the group labels in obtaining the scalar projections and therefore permutation tests on the T^2 statistic computed from these projections are essential in finding meaningful p-value maps. The permutation tests on a million voxels can be computationally expensive. To circumvent the permutation tests, we provide an alternative analytical kFDA solution, based upon eigen-decomposition, as is shown by Baudat et al. [7]. This analytic solution has been shown to be mathematically equivalent [45] to first performing kernel PCA on the input data, followed by ordinary FDA. Therefore, we shall refer to this alternative solution as kPCA-based kFDA. An advantage of the analytic solution is that one can reduce the number of kernel PCA components used in the subsequent ordinary FDA and obtain superior discriminatory performance. For good discriminatory performance, the number of kernel PCA components used in the subsequent ordinary FDA should not exceed the number of samples in either class. In practice, this number is chosen based upon the kernel PCA eigenvalue spectrum as discussed in Sect. 3. On account of the equivalence between the T^2 statistic and FDA (Appendix), a second advantage is that it is possible to compute p-values using the T^2 statistic on the retained kPCA components in a faster parametric manner, with a small loss in accuracy, in comparison to performing permutation tests on kFDA components. The p-value map computation procedure using kPCA-based kFDA is identical to Algorithm 2 presented in Sect. 3.

5 Application

The kPCA framework is applied on three types of datasets: (1) simulated 2D datasets with the purpose of testing parameters of the kPCA analysis; (2) real datasets in which changes in shape and orientation have been simulated to study the practical applicability of KBM to group-wise population studies. Knowing the ground truth, that is, the magnitude of changes introduced, makes it easier to evaluate the differences captured by kPCA analysis and identify possible false positives; and (3) A group analysis between children with Autism spectrum disorder (ASD) and typically developing (TD) controls. We now describe the details of each of the experiments.

5.1 Kernel Based Analysis of Simulated Datasets

In this experiment the aim is to establish that the kernel-based method is able to identify the changes in shape and orientation in tensors when the changes occur in combination as could be in the case of pathology-induced changes. A 2-dimensional dataset with variation in the radial and angular directions was created that modeled a tensorial dataset with changes in the principal eigenvalue and eigen-direction. The purpose for using only two-dimensions was to make understanding and visualization straightforward (Fig. 2).

The synthetic dataset consisted of points forming a semi-circular band (see Fig. 2a) and was generated using 36 angles (in the 0–144° range) and 6 radial values (in the range 1.3–1.8) resulting in a total of 216 points. The aim was to check whether kernel-based morphometry paradigm was able to capture both these changes. The kernel-based procedure (Algorithm 1) was applied to this dataset using a Gaussian radial basis function (RBF) with the kernel width σ^2 set to 0.1 ($\sigma = 0.316$). The kernel width parameter was based on the average distance between nearest neighbors $\mathbf{x}_i$ and $\mathbf{x}_j$, i.e. $||\mathbf{x}_i - \mathbf{x}_j||$ and the number of samples.

Figure 2c–i shows the iso-contour plots for 7 principal kPCA components representing the hyperplanes having constant projections onto the corresponding 7 RKHS eigenvectors, as was described earlier using Fig. 1. It was observed that the first 6 components (Fig. 2c–h) represented the angular changes in the data using varying scales, the third kPCA component (Fig. 2e) divided the angular variation in the data into four regions and alternately assumed positive and negative values as we move along the angular direction across these four regions. Only the seventh kPCA component (Fig. 2i) individually captured the radial change in the data and it smoothly increased from negative to positive values in the radial direction.

5.2 Analysis of DTI Datasets with Simulated Changes

The simulation in the previous Sect. 3 was then extended to a more realistic scenario as now the changes were simulated in real datasets. This was performed to determine whether KBM is able to capture combined shape and orientation changes, that a simple voxel-based morphometry of the scalar maps of anisotropy and diffusivity is unable to obtain.

The DTI data consisted of scans of 36 healthy volunteers (17 male and 19 female). These DT images were acquired on a Siemens Trio^{TM} 3.0 Tesla scanner, using a single shot spin-echo, echo planar imaging (EPI), with 12 diffusion directions with a b-value of 800 s/mm^2 and TR/TE = 6,400/97 ms. Forty axial slices with 128×128 matrix, were acquired with a voxel-size of $1.72 \times 1.72 \times 3.0$ mm. The diffusion tensor images were reconstructed from the DWI data using multivariate linear fitting [32]. The FA images that were computed from the tensors, were deformably registered elastically to a chosen healthy subject as template, by hierarchically matching features that provide a rich morphological signature for each voxel [36], The deformation was applied to the tensors while reorienting them using the underlying rotation component of the transformation [44]. We then identified an ROI on the template in the corpus callosum, as shown in Fig. 3a, and introduced spatially smooth random changes in the principal eigenvalue and the azimuthal angle for the principal eigenvector of each tensor into the appropriate ROI for all unwarped subject DT images. The random changes were designed to slightly increase the principal eigenvalue (average 4.6 % change) and the principal azimuthal angle ($\leq 20°$), on average, but were subtle enough so that these changes could not be visibly easily discerned on an FA map or a colormap for the principal direction. These changes emulated changes in FA and orientation in the tracts. The DT images with the introduced random changes were then warped back to the template resulting in 35 DT images belonging to the class with induced pathology. The tensors were then transformed to log-Euclidean space. KBM method was tested using two different cases: (1) using LE tensors (without smoothing) (2) by applying 4 mm FWHM Gaussian blur to the LE tensors. All the tests were performed using 3 kPCA components, 8 kPCA components and 12 kPCA components. Following kPCA, the statistical p-value maps were computed using Hotelling T^2 test. KBM method uses a vector (LE tensor in this case) as an input and hence can be applied for analyzing of one of the eigenvectors. To demonstrate this adaptability, we applied our method to analyze the orientation of the simulated tensors. The principal direction (PD) defined by a $3D$ vector was chosen for the analysis. The issue of antipodal symmetry of eigenvectors was resolved by making sure that all the vectors lay in the positive z-hemisphere. The kPCA framework described in Sect. 3 was applied to the PDs.

The effect of applying kPCA to PDs, tensors and the effect of change of parameters was quantitatively evaluated based on the percentage overlap of voxels

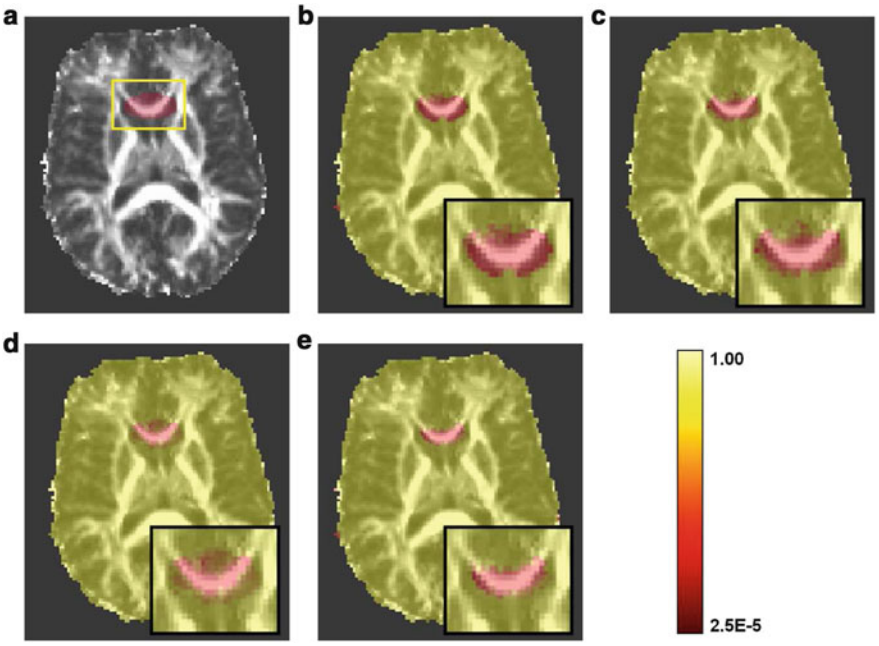


Fig. 3 (**a**) ROI with changes (highlighted) overlaid on the template FA map. P-value map computed from voxel-wise kPCA using 8 components is overlaid on the template FA when (**b**) Smooth tensors are used (**c**) when original tensors are used, (**d**) voxel-wise t-test on FA and (**e**) when principal eigen-directions are used. Low p-value regions in (**b**) better (≈66 %) match the true ROI in (**a**) than in (**c**) (≈59 %) after thresholding at p-value of 0.1. For FA analysis in (**d**), the sensitivity is lower than kPCA on tensors. Although most of the ROI is detected, the p-values are higher. For PD analysis in (**e**) changes in anisotropic areas are better detected than in isotropic regions. The ROI's are zoomed in each case for better visualization

in the detected ROI (the voxels that showed to be significantly different based on a threshold of p-value 0.1) with the voxels in the original ROI in which the changes were introduced. We compared our method with voxel-wise t-test on FA and ADC as well as the Isomap-based method introduced in [40]. The kPCA on PD's was compared with the Bipolar-Watson method on the PD's introduced by Schwartzman et al. in [34].

Results are shown in Fig. 3b–e. Figure 3b displays the p-value map after performing group analysis on the kernelized data at each voxel. It can be observed that the simulated ROI has a very low p-value range. Similarly, Fig. 3c shows the p-value map after performing kPCA analysis on tensors. Figure 3d, e show the ROI detected by t-test on FA and kPCA on PD's respectively. Table 1 gives all the percentage overlap values from kPCA compared to FA analysis, ADC analysis and Isomaps. The FA and ADC analysis involved a voxel-wise t-test.

Table 1 Percentage overlap of detected ROI (p-value map threshold at a cut-off of 0.1) with the ground-truth ROI in which changes had been introduced (please refer to Sect. 5)

	Analysis method	Original DTI	Smooth DTI
1.	ADC	14.92	21.71
2.	FA	39.14	42.39
3.	ISOMAP – 3 components	36.19	45.50
4.	kPCA on tensors – 3 components	57.39	65.56
5.	kPCA on tensors – 8 components	59.43	66.34
6.	kPCA on tensors- 12 components	64.38	87.42
7.	kPCA on PD – 3 components	42.15	45.76
8.	kPCA on PD – 8 components	42.87	47.55
9.	Bipolar Watson on PD	39.84	42.92

5.3 kPCA Analysis on Autism Spectrum Disorder

Finally, the method of KBM was applied to a real population of subjects with ASD pathology and typically developing (TD) controls. In this study 26 TD controls (mean age = 10.7) and 44 subjects with ASD (mean age = 9.8) were used. The images were taken using Siemens 3T Verio^{TM} scanner using a 32 channel head coil. DTI was performed using a single shot spin-echo, echo-planar sequence with the following parameters: TR/TE = 16,900/70 ms, b-value of 1,000 s/mm^2 and 30 gradient directions. Eighty axial slices of 128 $\times$ 128 matrix (FOV 256 mm) were acquired yielding 2 mm isotropic data. The diffusion tensors were estimated using the least squares fitting method and then spatially normalized to a standard template described in Wakana et al. [41]. The deformable registration utilized the full tensor information by integrating intensity and orientation into a hierarchical matching framework [22]. KBM analysis was then carried out on the tensors using an energy threshold of 80 % and with a sigma of 4.0. FA and mean diffusivity (MD) maps were computed from the spatially normalized dataset. A comparative voxel-wise FA and MD analysis was performed on the same dataset by employing a standard t-test between the groups. In all the analyses, the p-values were thresholded at a significance level of $p < 0.05$ and the results were overlaid on the template FA image. The resulting images are displayed in Fig. 4 indicating the regions of differences between subjects with ASD and the TD controls. The kernel based method could capture multiple areas of significance that included left superior longitudinal fasciculus (SLF), left inferior longitudinal fasciculus (ILF) and left inferior fronto-occipital fasciculus (IFO) and parts of right and left internal (IC) and external capsule (EC). The conventional FA and MD voxel-wise results are shown in Fig. 4b, c respectively. From Fig. 4b it can be observed that FA captures only the differences in the right external capsule (EC), while MD analysis shows significance only in inferior regions that include the ILF and IFO as seen in Fig. 4c. Multiple comparisons using FDR at 0.1 threshold (on tests for FA, MD and KBM on tensors)could not survive any of the voxels.

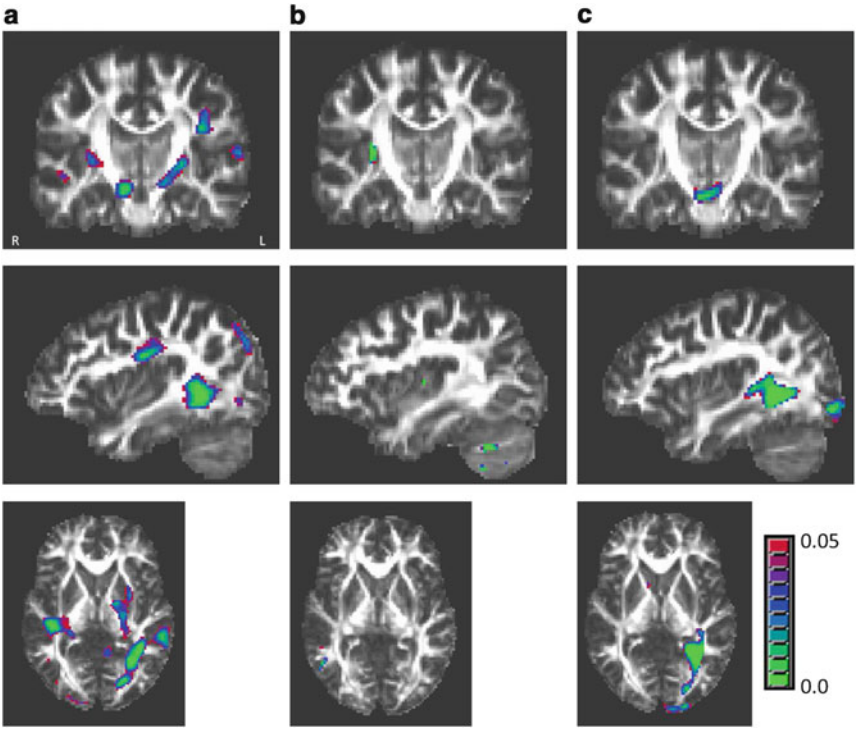


Fig. 4 Displays the differences between ASD and TD groups. (**a**) Result from k-PCA on tensors. The areas include left SLF (includes arcuate fasciculus and acoustic radiation), ILF, IFO, IC and EC while (**b**) show the result from voxel-wise FA analysis in which only the right EC is captured and (**c**) significant areas captured by MD changes that include ILF and IFO

6 Summary

In this chapter we address the problem of tensor-based population statistics of DTI data by employing a kernel based morphometric method that can capture the underlying distribution of the data. Our sequence of experiments shows that the mapping of data to a kernelized higher dimensional space enhances group separation and also models the underlying changes in the data.

To validate the kernel-based procedure we first applied it Sect. 5.1 on simulated data. This aided in determining whether it was able to capture combined changes in shape and orientation and whether this depended on the number of components. The experiment demonstrated that using different number of components achieves different degrees of separability in the data, between the different kinds of changes. Moreover, we found that it is important to utilize an adequate number of features/components for better group separation. Therefore, an energy threshold

criteria was defined as was described in Sect. 3. It was also noted that if a high value of σ was used (e.g. 10 times of NN distance) then the separability of the shape and orientation could not be obtained even when maximum allowable components were used. On the other hand, a low σ value would have lead to overfitting of the data. Thus selecting an optimum σ value was critical.

In the next experiment (Sect. 5.2) pathological changes were simulated in the genu of the corpus callosum and the surrounding CSF (in the form of spatially smooth subtle random changes in the principal eigenvalue and the PD). The main reason behind picking such an area was to evaluate KBM in areas with high anisotropy (genu) as well as with low anisotropy (CSF). These simulations were mainly created for better validation of the KBM method as there are no models or ground truth for the variability that a disease may introduce in the data. From the results shown in Table 1, rows 1 and 2 are the outcome of a conventional t-test on ADC and FA and row 3 presents the results using the Isomap technique from [40]. The t-test on FA detected only regions of higher percentage change in eigenvalues with a significance lower than 0.1. The Isomap technique (row 3) performed better than the first two approaches (indicating the non-linear nature of data variation), but perhaps it suffered as it does not utilize the knowledge of the underlying distribution of data. Knowledge of the statistical distribution led to improved results using the kPCA technique on the LE tensors (Table 1 rows 4–6). It was observed that the resulting overlap from kPCA was better when the LE tensors were smoothed and when larger number of features were used (87 % when 12 components were used on smoothed datasets). The results also improved when richer number of features were used (that is features had more components). In CSF, since the average eigenvalues were slightly increased, we expected to see shape differences which were significantly caught by kPCA method whereas FA picked it up subtly. Row 7 and 8 present results using kPCA on PD's while row 9 shows the results of using Bipolar-Watson model introduced in [34] to determine changes in PD. Although the ROI overlap using kPCA on PD's using 8 components, ($\approx$47 %) was much lower than the tensor overlap ($\approx$66 %), it was better than using the Bipolar-Watson method which showed only 42 % overlap after thresholding at a p-value of 0.1. Although the changes were introduced in the principal eigenvalue and eigenvector, it seems that methods that targeted each of these changes individually (FA-based and PD-based) were unable to capture it fully as opposed to when the full tensor information was used for statistical analysis. Thus suggests that changes in shape and orientation are difficult to detect by conventional methods, however the KBM is able to capture mixed changes. Since combination changes are expected in pathology, we have established the importance of our method in performing large population studies in which changes cannot be hypothesized a priori. It may be noted that if it were known a priori that the changes were only in shape or orientation, we could examine just that aspect, but in the absence of such knowledge, it is important to study the full tensor.

Finally, we performed KBM on a population of subjects with ASD (Sect. 5.3), to demonstrate the applicability of our method on clinical datasets. DTI based research in ASD has mainly involved studying WM changes using anisotropy and diffusivity values [39]. Abnormalities have been reported in WM structures like the genu and splenium of the corpus callosum [1], the internal capsule [23] and in the tempo-parietal regions [5, 21].

The results from our analysis displayed in Fig. 4 indicate the regions of differences between subjects with ASD and the TD controls. The resulting differences from the kernel-based method suggest WM abnormalities and hypo-connectivity between brain regions which may be strong contributors to the social deficits that are hallmarks of the ASD phenotype. For example, the changes observed in SLF (which includes the arcuate fasciculus) can be linked to the language impairment often observed in the ASD population [16] while the differences in the internal capsule were comparable to the previous finding by Keller et al. [23]. The p-values computed from the kernelized analysis as well from FA and MD analyses could not survive the FDR correction (at 0.1 threshold), perhaps owing to the heterogeneity in the population, small sample size and/or subtle differences between the groups. However, the aim here was to demonstrate future clinical applicability of the method as it was able to capture more changes than were observed using conventional analysis of DTI.

It was shown in Sect. 5.2, that tensor analysis could capture interplay between combined shape and orientation changes which individual (FA or PD) analysis could not capture. Similarly in the ASD example, the significant regions using kPCA on tensors in Fig. 4a–c included many areas like SLF and IC that are known to be affected in ASD while FA failed to capture the pathological abnormalities in ASD. This demonstrates that KBM of tensors is able to find combined FA and PD changes that other methods are unable to, underlining the importance of full tensor statistics, in comparison to statistics on the scalar maps alone.

In summary, DTI analysis has the advantage of being more sensitive than the standard scalar or PD analyses, especially when the changes appear in combination (that is shape and orientation as could be the case in real data). However more work needs to be done for interpretation of results which will vary based on the dataset on which KBM has been applied. However, this work establishes the need for full tensor statistics in group-wise population studies. As the effect of pathology is not known, tensor analysis can be thought of as an unbiased method rather than using scalar indices computed from combination of tensor eigenvalues and eigenvectors. The wide range of experiments demonstrate applicability of kernel-based tensor morphometry for population statistics and provides a novel method of statistical analysis, based on capturing the underlying distribution of the data, that is specific to the disease that introduced the change.

Acknowledgements This work was supported by the NIH grants RO1-MH079938 and R01-MH092862.

Appendix

Proof of the equivalence between Hotelling's T^2 test and FDA [24].

Let $\mathbf{x}_{1i}, i = 1, \cdots, N_1$ be the p-dimensional data vectors belonging to class 1 and let $\mathbf{x}_{2i}, i = 1, \cdots, N_2$ be the p-dimensional data vectors belonging to class 2. Let $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$ denote the means for classes 1 and 2. Let $S_x = (\frac{1}{N_1+N_2-2})\Big(\sum_{i=1}^{N_1}(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)^T + \sum_{i=1}^{N_2}(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)^T\Big)$.

Hotelling's T^2 test: Then the T^2 test statistic is:

$$T_x^2 = \frac{N_1 N_2}{N_1 + N_2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \tag{3}$$

Assuming normal distributions for $\mathbf{x}_{1i}, i = 1, \cdots, N_1$ and $\mathbf{x}_{2i}, i = 1, \cdots, N_2$ implies that $F_x = \frac{N_1+N_2-p-1}{(N_1+N_2-2)p} T_x^2$ has the cdf $F(p, N_1 + N_2 - p - 1)$. Therefore, the parametric form of Hotelling's T^2 test is sometimes refered to as the F-test.

FDA: The FDA optimal linear discriminant direction is $w = S_x^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ and the corresponding scalar mapping is $y = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1} x$.

Therefore, $\bar{y}_1 - \bar{y}_2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$,

$$\begin{aligned} S_y &= (\frac{1}{N_1 + N_2 - 2})\Big(\sum_{i=1}^{N_1}(y_{1i} - \bar{y}_1)(y_{1i} - \bar{y}_1)^T + \sum_{i=1}^{N_2}(y_{2i} - \bar{y}_2)(y_{2i} - \bar{y}_2)^T\Big) \\ &= (\frac{1}{N_1 + N_2 - 2})(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1}\Big(\sum_{i=1}^{N_1}(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)^T \\ &\quad + \sum_{i=1}^{N_2}(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)^T\Big) S_x^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \qquad (4) \\ &= (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \qquad (5) \end{aligned}$$

and

$$\begin{aligned} T_y^2 &= \frac{N_1 N_2}{N_1 + N_2}(\bar{y}_1 - \bar{y}_2)^T S_y^{-1}(\bar{y}_1 - \bar{y}_2) \\ &= \frac{N_1 N_2}{N_1 + N_2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T S_x^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \\ &= T_x^2 \end{aligned} \tag{6}$$

Thus, the T^2 statistic computed on vectorial input samples, i.e. T_x^2, is mathematically equivalent to a one-dimensional T^2 statistic T_y^2 computed on the scalar samples obtained by performing Fisher discriminant analysis on the input vectorial samples.

References

1. Alexander, A.L., Lee, J.E., Lazar, M., Boudos, R., DuBray, M.B., Oakes, T.R., Miller, J.N., Lu, J., Jeong, E.K., McMahon, W.M., Bigler, E.D., Lainhart, J.E.: Diffusion tensor imaging of the corpus callosum in autism. Neuroimage **34**, 61–73 (2007)
2. Ardekani, B.A., Tabesh, A., Sevy, S., Robinson, D.G., Bilder, R.M., Szeszko, P.R.: Diffusion tensor imaging reliably differentiates patients with schizophrenia from healthy volunteers. Hum. Brain. Mapp. **32**, 1–9 (2011)
3. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magn. Reson. Med. **56**, 411–421 (2006)
4. Ashburner, J., Friston, K.J.: Voxel-based morphometry–the methods. Neuroimage **11**, 805–821 (2000)
5. Barnea-Goraly, N., Kwon, H., Menon, V., Eliez, S., Lotspeich, L., Reiss, A.L.: White matter structure in autism: preliminary evidence from diffusion tensor imaging. Biol. Psychiatry **55**, 323–326 (2004)
6. Basser, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. J. Magn. Reson. B **111**, 209–219 (1996)
7. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. Neural Comput. **12**, 2385–2404 (2000)
8. Behrens, T.E.J., Berg, H.J., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage **34**, 144–155 (2007)
9. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., Golani, I.: Controlling the false discovery rate in behavior genetics research. Behav. Brain. Res. **125**, 279–284 (2001)
10. Burges, C.: Data mining and knowledge discovery handbook. Kluwer, New York (2005)
11. Calamante, F., Masterton, R.A.J., Tournier, J.D., Smith, R.E., Willats, L., Raffelt, D., Connelly, A.: Track-weighted functional connectivity (TW-FC): a tool for characterizing the structural-functional connections in the brain. Neuroimage **70**, 199–210 (2013)
12. Cao, Y., Miller, M., Mori, S., Winslow, R.: Diffeomorphic matching of diffusion tensor images. In: Proceedings of CVPR-MMBIA, New York (2006)
13. Field, A.S., Alexander, A.L.: Diffusion tensor imaging in cerebral tumor diagnosis and therapy. Top. Magn. Reson. Imaging **15**, 315–324 (2004)
14. Fletcher, T., Joshi, S.: Riemannian geometry for the statistical analysis of diffusion tensor data. Signal Process. **87**(2), 250–262 (2007)
15. Fletcher, T., Lu, C., Pizer, S., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans. Med. Imaging **23**, 995–1005 (2004)
16. Fletcher, P.T., Whitaker, R.T., Tao, R., DuBray, M.B., Froehlich, A., Ravichandran, C., Alexander, A.L., Bigler, E.D., Lange, N., Lainhart, J.E.: Microstructural connectivity of the arcuate fasciculus in adolescents with high-functioning autism. Neuroimage **51**, 1117–1125 (2010)
17. Genovese, C.R., Lazar, N.A., Nichols, T.: Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage **15**, 870–878 (2002)
18. Girolami, M.: Orthogonal series density estimation and the kernel eigenvalue problem. Neural Comput. **14**, 669–688 (2002)
19. Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O.: Mapping the structural core of human cerebral cortex. PLoS Biol **6**, e159 (2008)
20. Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V.J., Meuli, R., Thiran, J.P., Grant, P.E.: White matter maturation reshapes structural connectivity in the late developing human brain. Proc. Natl. Acad. Sci. U S A **107**, 19067–19072 (2010)
21. Ingalhalikar, M., Parker, D., Bloy, L., Roberts, T.P.L., Verma, R.: Diffusion based abnormality markers of pathology: toward learned diagnostic prediction of ASD. Neuroimage **57**, 918–927 (2011)

22. Ingalhalikar, M., Yang, J., Davatzikos, C., Verma, R.: DTI-DROID: Diffusion tensor imaging-deformable registration using orientation and intensity descriptors. Int. J. Imaging Syst. Technol. **20**(2), 99–107 (2010)
23. Keller, T.A., Kana, R.K., Just, M.A.: A developmental study of the structural integrity of white matter in autism. Neuroreport **18**, 23–27 (2007)
24. Khurd, P., Verma, R., Davatzikos, C.: Kernel-based manifold learning for statistical analysis of diffusion tensor images. Inf. Process. Med. Imaging **20**, 581–593 (2007)
25. Kubicki, M., Westin, C.F., McCarley, R.W., Shenton, M.E.: The application of DTI to investigate white matter abnormalities in schizophrenia. Ann. N Y Acad. Sci. **1064**, 134–148 (2005)
26. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing. J. Math. Imaging Vis. **25**(3), 423–444 (2006)
27. Li, Y., Zhu, H., Y, C., Ibrahim, J., An, H., Lin, W., Hall, C., Shen, D.: Radti: regression analyses of diffusion tensor images. In: SPIE Conference Proceedings, Dresden (2009)
28. Lim, K.O., Hedehus, M., Moseley, M., de Crespigny, A., Sullivan, E.V., Pfefferbaum, A.: Compromised white matter tract integrity in schizophrenia inferred from diffusion tensor imaging. Arch. Gen. Psychiatry **56**, 367–374 (1999)
29. Nimsky, C., Ganslandt, O., Hastreiter, P., Wang, R., Benner, T., Sorensen, A.G., Fahlbusch, R.: Preoperative and intraoperative diffusion tensor imaging-based fiber tracking in glioma surgery. Neurosurgery **56**, 130–7 (2005). Discussion 138
30. Park, H.J., Westin, C.F., Kubicki, M., Maier, S.E., Niznikiewicz, M., Baer, A., Frumin, M., Kikinis, R., Jolesz, F.A., McCarley, R.W., Shenton, M.E.: White matter hemisphere asymmetries in healthy subjects and in schizophrenia: a diffusion tensor MRI study. Neuroimage **23**, 213–223 (2004)
31. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. Int. J. Comput. Vis. **66**(1), 41–66 (2006)
32. Pierpaoli, C., Jezzard, P., Basser, P., Barnett, A.: Diffusion tensor MR imaging of the human brain. Radiology **201**, 637 (1996)
33. Scholkopf, B., Smola, A.: Learning with Kernels. MIT, Cambridge (2002)
34. Schwartzman, A., Dougherty, R.F., Taylor, J.E.: Cross-subject comparison of principal diffusion direction maps. Magn. Reson. Med. **53**, 1423–1431 (2005)
35. Schwartzman, A., Dougherty, R.F., Taylor, J.E.: Group comparison of eigenvalues and eigenvectors of diffusion tensors. J. Am. Stat. Assoc. **105**(490), 588–599 (2010)
36. Shen, D., Davatzikos, C.: HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging **21**, 1421–1439 (2002)
37. Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.J.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage **31**, 1487–1505 (2006)
38. van den Heuvel, M.P., Mandl, R.C.W., Kahn, R.S., Hulshoff Pol, H.E.: Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. Hum. Brain Mapp. **30**, 3127–3141 (2009)
39. Verhoeven, J.S., Cock, P.D., Lagae, L., Sunaert, S.: Neuroimaging autism. Neuroradiology **52**, 3–14 (2010)
40. Verma, R., Khurd, P., Davatzikos, C.: On analyzing diffusion tensor images by identifying manifold structure using isomaps. IEEE Trans. Med. Imaging **26**, 772–778 (2007)
41. Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C.M., Mori, S.: Fiber tract-based atlas of human white matter anatomy. Radiology **230**, 77–87 (2004). http://radiology.rsnajnls.org/cgi/reprint/230/1/77.pdf.
42. Wu, Y.C., Field, A.S., Chung, M.K., Badie, B., Alexander, A.L.: Quantitative analysis of diffusion tensor orientation: theoretical framework. Magn. Reson. Med. **52**, 1146–1155 (2004)
43. Xu, D., Hao, X., Bansal, R., Plessen, K.J., Peterson, B.S.: Seamless warping of diffusion tensor fields. IEEE Trans. Med. Imaging **27**, 285–299 (2008)

44. Xu, D., Mori, S., Shen, D., van Zijl, P.C.M., Davatzikos, C.: Spatial normalization of diffusion tensor fields. Magn. Reson. Med. **50**, 175–182 (2003)
45. Yang, J., Frangi, A.F., Yang, J.Y., Zhang, D., Jin, Z.: KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 230–244 (2005)
46. Zhang, H., Yushkevich, P.A., Alexander, D.C., Gee, J.C.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. Med. Image Anal. **10**, 764–785 (2006)

The Estimation of Free-Water Corrected Diffusion Tensors

Ofer Pasternak, Klaus Maier-Hein, Christian Baumgartner, Martha E. Shenton, Yogesh Rathi, and Carl-Fredrik Westin

Abstract Diffusion tensor imaging (DTI) is sensitive to micron scale displacement of water molecules, providing unique insight into microstructural tissue architecture. The tensors provide a compact way to describe the average of these displacements that occur within a voxel. However, current practical image resolution is in the millimeter scale, and thus diffusivities from many tissue compartments are averaged in each voxel, reducing the specificity of the measurement to subtle pathologies. In this chapter we review the free-water model, and use it to derive diffusion tensors following the elimination of the free-water component, that is assumed to originate from the extracellular space. Doing so, the resulting diffusion tensors and their derived indices measure the tissue itself, and are more sensitive to the geometry of the tissue, increasing the specificity to pathologies that affect brain tissue.

1 Introduction

Diffusion tensor imaging (DTI) along with other diffusion imaging analysis methods have became the leading MRI methodology to investigate the microstructure of brain tissue. In almost two decades of research since its inception [1], DTI studies

O. Pasternak (✉) • Y. Rathi • C.-F. Westin
Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
e-mail: ofer@bwh.harvard.edu

K. Maier-Hein
German Cancer Research Center, Heidelberg, Germany

C. Baumgartner
King's College London, London, UK

M.E. Shenton
Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

VA Boston Healthcare System, Brockton, MA, USA

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_11, © Springer-Verlag Berlin Heidelberg 2014

were able to detect abnormalities that occur due to various brain disorders that include stroke, traumatic brain injuries, Multiple Sclerosis, Alzheimer's disease, Parkinson, Schizophrenia and many more disorders [2]. It appears that DTI indices are very sensitive to even subtle changes, either in the neuronal tissue or its surrounding, implicating a variety of pathologies, including demyelination, vasogenic and cytotoxic edema, inflammation, cell swelling, gliosis and other changes in the shape or sizes of neuron and glia cells [3]. Since DTI is very sensitive to many different pathologies, the challenge is to find methods that increase the specificity of the DTI indices, in such a way that allows better differentiation and diagnosis of the underlying processes that lead to the disorder, symptoms or pathology.

Diffusion MRI (dMRI) measures the displacement of water molecules, which in a typical brain imaging experiment displace a few tens of microns. This makes dMRI sensitive to normal and pathological architecture in the cellular scale. Nevertheless, current image resolution is in the millimeter scale, introducing partial volume of different tissue types – white matter, gray matter, glia cells, cerebrospinal fluid (CSF) – which reduces the sensitivity and specificity of most indices derived from dMRI and DTI [4]. Controlling for partial volume, and models that account for multiple compartments can therefore help in increasing the specificity, by determining in which of the compartments the abnormality occurs.

To date, most partial volume elimination methods concentrate on the case where two or more white matter fibers share a voxel (e.g., [5–7]). This chapter concentrates on the partial volume that occurs between the intra- and extra-cellular compartments. Correcting for extracellular water is required to eliminate CSF contamination, thus improving DTI's sensitivity in the vicinity of the ventricles [8] and important for the delineation of fibers that pass next to the CSF, such as the fornix [9, 10]. Moreover, the fractional volume of the extracellular water, relative to the remaining hindered or restricted water molecules, appears to provide important information with regard to pathological processes that modify the interstitial extracellular space, such as edema [9], neuroinflammation [11] and atrophy [12]. Indeed, the extracellular volume was shown to be sensitive to pathologies that appear in aging [12], schizophrenia [13], multiple-sclerosis [11], and Alzheimer's disease [14].

In this chapter we describe methods to estimate diffusion tensors that represent the signal following the elimination of free-water in the extracellular space. We start in Sect. 2 by describing the free-water model. We then show in Sect. 3 how to estimate the model parameters from a conventional, single-shell DTI acquisition, and from a more sophisticated multi-shell dMRI acquisition. In Sect. 4 we describe an alternative approach that estimates the model parameters in white matter while performing tractography. We conclude in Sect. 5 by demonstrating the effect of using free-water along with a clustering method to perform a group comparison between Alzheimer's disease patients and age matched controls.

2 The Free-Water Model

Free-water is defined as self-diffusing water molecules that do not experience restriction or hindrance from their surrounding. This definition depends on the amount of time in which the diffusion process is measured, as for infinitely long diffusion times, all water molecules will eventually hit an obstacle, whereas for infinitely short diffusion time, all water molecules are free to diffuse. In typical experimental setups the diffusion time is in the order of 30–50 ms which yields that in order for water molecules to diffuse freely they have to be a few tens of microns away from membranes and other cellular restrictions. The typical size of brain cells is usually lower than 10 μm, and therefore in the brain, free-water can be found in the extra-cellular space. The diffusion coefficient of free-water is known, and depends on the temperature. For example, the diffusion coefficient of water in body temperature is $d_{water} = 3 \times 10^{-3}\,\mathrm{mm}^2/\mathrm{s}$ [15].

The free water model was first proposed by Pierpaoli and Jones [16]. The model estimates and corrects for the contribution of free-water [9]. The model assumes that the diffusion signal originates from two molecular compartments, co-existing within a voxel, with slow exchange between the compartments [9]:

$$\hat{A}_i(\mathbf{D}, f) = f \exp(-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i) + (1 - f) \exp(-b_i d)\,. \tag{1}$$

Here, $\hat{A}_i$ is the estimated signal (normalized by the b_0) of the i'th applied diffusion gradient with orientation $\mathbf{g}_i$, and b-value b_i. The first term reflects the tissue compartment, where $\mathbf{D}$ is the diffusion tensor of this compartment and f is the relative contribution of the compartment. The second term reflects an isotropic compartment, with a fixed diffusion coefficient, d, set to the diffusion coefficient of water in body temperature, $3 \times 10^{-3}\,\mathrm{mm}^2/\mathrm{s}$. Thus, the isotropic compartment models free-water, and is expected to measure molecules that are in the extra-cellular space, we call this compartment the free-water compartment. If the free-water compartment is eliminated, then the remaining compartment measures molecules that are close to tissue membranes, these are expected to include all the intracellular molecules, and some of the extracellular molecules that are not far enough from hindering obstacles such as cellular membranes.

When the free-water model was introduced it was different than other bi-compartmental models by the fact that the diffusion coefficient of the isotropic compartment was fixed. Prior to the free-water, other bi-compartment models either did not restrict the diffusivities of the compartments, resulting with "fast and slow" components (see for example [17]), or extremely restricted tissue compartment, with models such as the "ball-and-stick" [18] that separated all of the isotropic contribution (ball) from any remaining anisotropic contribution (stick). However, it was not clear how to relate the resulting compartments to biological compartments: In the fast-and-slow model, the fitting did not seem to match known values of intra- and extra-cellular volumes, leading to a debate on what are the sources of these

compartments [17]. With the "ball-and-stick" model, the orientation of the stick is generally aligned with that of white matter bundles, however the model does not allow assessment of other diffusion properties of the white matter, limiting the usability of this model to tractography studies. The fixed diffusivity of the free-water compartment, provides more of a biological compartmentalization to extra-cellular versus tissue compartments.

2.1 Free-Water Derived Maps

To acquire free-water maps, the free-water model in Eq. (1) has to be fitted with the measured dMRI signal. The outcome of this model-fit is a map of the tissue fraction (f) or a free-water map ($1 - f$). The free-water map, for example, has a range of [0–1] where values close to 0 are expected to be found in densely packed tissue, such as major fiber bundles, and values close to 1 are expected to be found in areas filled with CSF, such as the ventricles. In addition, the free-water eliminated tensor **D** can be further decomposed into scalar indices that can also be visualized as maps. These quantities include Fractional Anisotropy (FA), Mean diffusivity (MD), Axial and Radial diffusivity [19]. These tensor quantities describe properties of the tissue compartment, as such, the FA derived from the tissue compartment is called tissue-FA or FA_t. This notation differentiates the value from the FA that is derived from DTI, reflecting the anisotropy in the entire voxel (i.e., both the tissue and free-water compartments).

3 Fitting the Free-Water Model

The free-water model adds only one more parameter, f, to the DTI model. However, unlike DTI, the fitting of this bi-exponential model is highly unstable [17]. In order to stabilize the fit, we need to incorporate additional information that reduces the number of possible solutions.

3.1 Single Shell (DTI) Data

One way to stabilize a fit is by regularizing the fitted function. This is since requiring from a solution to be piece-wise smooth restricts considerably the solution space. Regularization is performed by including neighborhood information in the fitting of each voxel. This information requires that the solution not only fit the measurement, but will also be similar enough to the solution of the neighboring voxels. To stabilize the fitting process we use a regularization framework that adds

neighborhood constraints as additional information. We use a regularization method based on the Beltrami operator [20] by minimizing the following functional:

$$L(\mathbf{D}, f) = \int_{\Omega} \sum_{i \in G} ||\hat{A}_i - A_i|| + \alpha \sqrt{|\gamma(\mathbf{D})|}\,. \tag{2}$$

Here, Ω includes all voxels of interest, G are the indexes of all applied gradients and A_i are their signal normalized by the b_0. The parameter α scales the contribution of the Polyakov action regularization term (typically $\alpha = 1$ [9]), with $|\gamma(\mathbf{D})|$ as the determinant of the induced metric. Using the Einstein summation convention this metric has the form $\gamma_{\mu\nu}(\mathbf{p}) = \partial_\mu \mathbf{x}^m \partial_\nu \mathbf{x}^n h_{mn}(\mathbf{x})$, for each coordinate in the space: $\mathbf{p} = \{x, y, z\}$.

In order to define the spatial-feature metric $\mathbf{H} = \{h_{mn}\}$, one has to define the spatial, and the feature metrics. Selecting a Euclidean metric with the canonical tensor representation for the feature metric (distances between tensors), simplifies the analysis considerably, and is also preferred over other types of global tensor metrics [21]. As a result the vector $\mathbf{x}$ has the elements

$$\mathbf{x} = [D_{xx}, D_{yy}, D_{zz}, \sqrt{2}D_{xy}, \sqrt{2}D_{yz}, \sqrt{2}D_{xz}, x, y, z]\,.$$

The metric $\mathbf{H}$ for a Euclidean tensor space is simply a 9×9 diagonal matrix, with 1 for the last 3 diagonal entries (the spatial domain) and a constant for the remaining 6 diagonal entries. This leads to the motion equations for the six tensor elements, $\mathbf{x}^j$ with $j \in \{1, 2, \ldots, 6\}$:

$$\Delta \mathbf{x}^j = \sum_{i \in G} b_i (\hat{A}_i - A_i) \exp(-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i) \left(\mathbf{g}_i^T \frac{\partial \mathbf{D}}{\partial \mathbf{x}^j} \mathbf{g}_i \right) + \frac{\alpha}{\sqrt{|\gamma|}} \partial_\mu \sqrt{|\gamma|} (\gamma^{\mu\nu} \partial_\nu \mathbf{x}^j)\,, \tag{3}$$

and for the fractional volume parameter:

$$\Delta f = \sum_{i \in G} -b_i (\hat{A}_i - A_i) \left(\exp(-b_i \mathbf{g}_i^T \mathbf{D}_t \mathbf{g}_i) - \exp(-b_i d) \right)\,. \tag{4}$$

The parameter f is maintained in the range $f \in [0, 1]$ by projecting values that exceed this range back within the range; see [9] for further constraints that can be enforced on this parameter. Importantly, due to the use of the Euclidean metric, and unlike the motion equations derived in [9], Eq. (3) does not have any Christoffel numbers, and therefore its calculation is simpler and faster. The second term in Eq. (3) is the Laplace-Beltrami operator, which is a piece-wise smooth, edge preserving tensor regularization operator [9, 20]. The final result is thus the parameters f and $\mathbf{D}$ that best fit the data while maintaining continuous tissue representation.

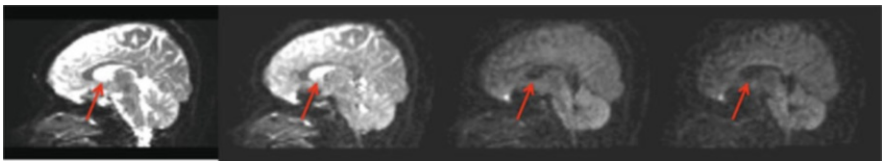


Fig. 1 Multi-shell acquisition. Diffusion signal for increasing b-values of (left to right) 0, 200, 900 and 1,400 mm^2/s in a mid-sagittal plane. The signal from free-water, such as in the ventricles (*red arrow*), decays into the noise floor faster than the signal of brain tissue

3.2 Multi-shell Free-Water Estimation

Regular DTI data is acquired using a single b-value, usually in the order of 1,000 s/mm^2, along with a non-diffusion-weighted image ($b = 0$ or b_0). The diffusion images are acquired using different gradient orientations, constituting a shell [22]. More elaborate acquisition schemes are available in which a number of different b-values are acquired, and a shell is acquired for each b-value, hence forming a multi-shell acquisition. Estimating the free-water fraction, f and the free-water eliminated tensor $\mathbf{D}$ using multi-shell data can be done using the same minimization defined in Eq. 2. However, when having multi-shell data, special properties of the multi-shell information can help initializing the estimation much closer to the minima [23]. When the initialization is good enough, the model fit may not require any further minimization.

The diffusivities of white and gray matter are considerably lower than those of free-water or CSF. Typically, in single shell DTI, healthy brain tissue has a quite homogeneous mean diffusivity of around 0.8 mm^2/s, 3–4 times slower than free-water. Therefore, the free-water signal is expected to decay faster than tissue, e.g., with a b-value of 900 s/mm^2 the tissue decays to 49 % of the signal while free-water decays to 7 % of the signal. Figure 1 shows an example of a multi-shell acquisition for a range of b-values. This range is achieved by modifying the diffusion gradient amplitude for fixed diffusion times. As expected, the free-water signal (mainly seen in the ventricles and around the parenchyma) attenuates faster than other brain tissue. The free-water signal diminishes completely into the noise floor for the higher b-values.

We calculate $\mathbf{D}_H$, the apparent diffusion tensor for the high b-valued shells, as an estimator for $\mathbf{D}$ by minimizing:

$$\sum_{i \in G_H} ||\hat{b_0} \exp(-b_i \mathbf{g}_i^T \mathbf{D}_H \mathbf{g}_i) - E_i|| \,, \tag{5}$$

where G_H are the indexes of all the applied gradients within the high b-valued shells, and E_i is the signal of the i'th acquired diffusion image, not normalized by the acquired b_0 (as opposed to its attenuation $A_i = E_i / E_0$). The acquired b_0 reflects the contribution of all spins within the voxel, including from free-water. Therefore,

the estimation of $\mathbf{D}_H$ requires the estimation of $\hat{b_0}$, which is the baseline image that would have been acquired in the case that the tensor $\mathbf{D}_H$ was the only component in the voxel. We minimize Eq. (5) using a linear least square (LLS) approach with $ln(\hat{b_0})$ as one of the free parameters [8].

We estimate f, which reflects the extracellular relative volume in a voxel, using the low b-valued shells, which are in the range that still has signal from free-water. Given $\mathbf{D}_H$ as an estimate for $\mathbf{D}$, we can calculate f_L as an estimate for f using LLS by defining:

$$f_L = (\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T\mathbf{c}\,, \tag{6}$$

where $\mathbf{c}_i = A_i - \exp(-b_i d)$ and $\mathbf{a}_i = \exp(-b_i \mathbf{g}_i^T \mathbf{D}_H \mathbf{g}_i) - \exp(-b_i d)$, and $i \in G_L$ being the indexes of the applied gradients in the low b-valued shells. Unlike our approach here, the single-shell free-water map estimation is initialized by the b_0 image alone, normalized by baseline values that assumed knowledge of voxels that have no tissue, and voxels that have no free-water [9]. This implicitly assumes that the T2 weighted images behave similarly across the entire brain, and that there are such baseline voxels. These assumptions are no longer required if using f_L and $\mathbf{D}_H$ as initialization.

3.3 Single- and Multi-shell Comparison

We test the multi-shell estimation on an acquisition optimized for the free-water estimation, having a single b = 0, 3 × b = 50, 6 × b = 200, 10 × b = 500, 30 × b = 900 and 16 × b = 1,400, with gradient orientations designed as nested platonic solids, which means that each shell is rotationally invariant, and the shells complement each other to a rotationally invariant scheme [24]. Data was acquired on a 1.5 T scanner with 2.5 mm isotropic voxels and takes 9:20 min. We use the b = 1,400 and b = 900 shells to estimate $\mathbf{D}_H$, and the remaining shells to estimate f_L.

All data was motion and eddy-current corrected. We used 3D-Slicer's tensor estimation to calculate $\mathbf{D}_H$ by first omitting all images in G_L, including the b_0. We used Matlab (Natick, MA) to calculate f_L. The complete analysis for a whole brain takes less than 15 min on a 64-bit Linux machine with Xeon-E5530 processors, without taking advantage of multiple cores.

Both the multi-shell (Fig. 2a) and single-shell (using the b = 900 shell; Fig. 2b) estimations provide similar free-water maps, showing high values in the ventricles, and low values in the brain tissue, nicely depicting the extracellular volume. The color by orientation maps are similar as well. To better evaluate the differences between the maps, we plot free-water maps using a color-scale that increases the visibility of the lower values. We can then see that the multi-shell maps are not as smoothed as the single-shell maps, although both estimations used the same Laplace-Beltrami regularization operator. As a result the multi-shell map is

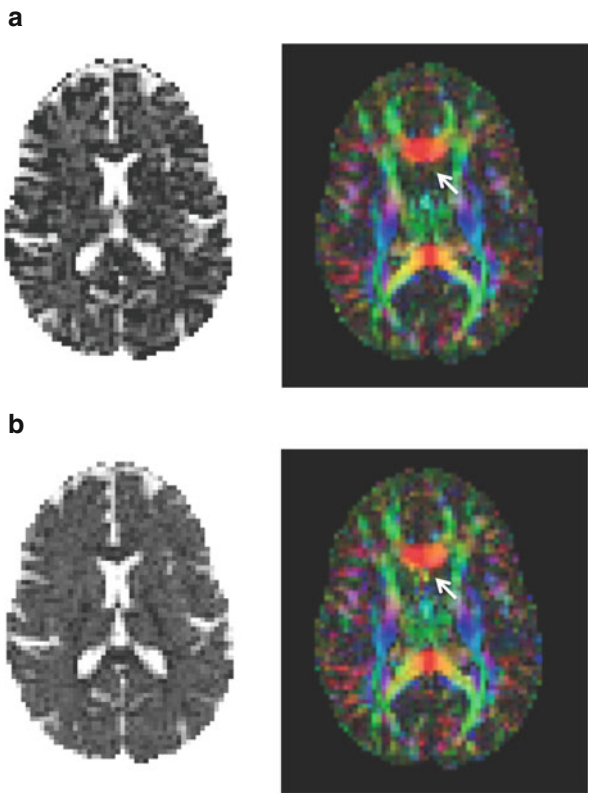


Fig. 2 Optimized multi-shell. Multi-shell regularized fitting (**a**) provides a more detailed extracellular volume (*left*) and color by orientation tissue tensor maps (*right*), comparing with the regularized single-shell fitting (**b**). Small details are preserved and there are less artifacts (*white arrows*)

more detailed, allowing to better distinguish cortical structures. For this acquisition scheme, the initialization, f_L is very similar to the final multi-shell free-water map.

4 Estimation via Filtered Tractography

An alternative to the gradient descent scheme proposed in the previous sections is to perform filtered tractography [6] to estimate the free-water volume along fiber tracts. The filtered tractography views tractography as a causal process and incorporates information from neighboring voxels to aid in the model fit [25]. We arrive at each new position along the fiber based upon the model parameters found at the previous position. As we examine the signal at each new position, a Kalman filter recursively updates the underlying local model parameters based on the last state, provides the variance of that estimate, and indicates the direction in which to propagate tractography, where the estimation begins again. Recursive estimation in this manner improves accuracy resolving individual orientations and yields inherently smooth tracts despite the presence of noise and uncertainty.

4.1 State-Space Representation

To begin estimating within a finite dimensional state-space filter we need the method-specific definition of four filter components:

1. The system state **x**: the model parameters
2. The state transition $\mathscr{F}[\cdot]$: how the model changes as we trace the fiber
3. The observation $v[\cdot]$: how the signal appears given a particular state
4. The measurement **y**: the signal we are trying to fit the model to

Similar to the fitting of Eq. (2), the model parameters consist of the tensor of the tissue compartment, **D**, and the fractional volume of the compartments, f. However, since the filtered tractography approach is restricted to following white matter, we can further simplify the model with the assumption that **D** is a cylindrical symmetric tensor, i.e., replace it with two eigenvalues, λ_1 and λ_2, and the principal direction **m**. The system state is therefore:

$$\mathbf{x} = [\mathbf{m}, \lambda_1, \lambda_2, f]^T \ ,$$

Typically, the local fiber configuration does not undergo drastic change from one position to the next. For this reason we assume identity dynamics for the state transition function. The observation function is the reconstruction of the attenuated signal, given by the free-water model in Eq. (1), i.e.,

$$v[\mathbf{x}] = \hat{A}_1, \ldots, \hat{A}_k \ ,$$

for k applied gradients. The measurement is the actual attenuated signal:

$$\mathbf{y} = A_1, \ldots, A_k$$

interpolated directly from the diffusion weighted images at the current position.

4.2 Unscented Kalman Filter

Since the relation between the signal and the free-water model parameters are non-linear, we employ an unscented Kalman filter to perform the estimation [6]. Similar to a classical linear Kalman filtering, the unscented version seeks to reconcile the predicted state of the system with the measured signal. The filter process consists of two stages: first the system transition model is used to predict the next state and observation, then the new measurement is used to correct the state estimate.

The Kalman filter is a particle filter that for a solution space with a dimension n requires $2n + 1$ noisy realizations, in our case, $n = 6$. The Prediction begins with the formation of a set $\mathbf{X}_t = \{\chi_i\} \subset \mathbb{R}^n$ of $2n + 1$ sigma point states (or particles)

with associated convex weights, $w_i \in \mathbb{R}$. The sigma points are spread around the current state using a Gaussian distribution, with mean, $\mathbf{x}_t \in \mathbb{R}^n$, and covariance, $\mathbf{P}_t \in \mathbb{R}^{n \times n}$. The sigma points and their weights are defined as follows:

$$\begin{aligned} \chi_0 &= \mathbf{x}_t \\ w_0 = \frac{\kappa}{(n+\kappa)} \quad w_i &= w_{i+n} = \frac{1}{2(n+\kappa)} \\ \chi_i = \mathbf{x}_t + \left[\sqrt{(n+\kappa)\mathbf{P}_t}\right]_i \quad \chi_{i+n} &= \mathbf{x}_t - \left[\sqrt{(n+\kappa)\mathbf{P}_t}\right]_i \end{aligned} \tag{7}$$

where $[\cdot]_i$ denotes the i-th column of a matrix and κ is an adjustable scaling parameter. We used $\kappa = 0.01$ in all our experiments. Next, this set of sigma points is propagated through the state transition function, $\hat{\chi} = \mathscr{F}[\chi] \in \mathbb{R}^n$, to obtain a new predicted sigma point set: $\mathbf{X}_{t+1|t} = \{\mathscr{F}[\chi_i]\} = \{\hat{\chi}_i\}$. As mentioned above, we assume identity dynamics, i.e., $\mathbf{X}_{t+1|t} = \mathbf{X}_t$. The predicted system state is then calculated as the weighted average of the set,

$$\bar{\mathbf{x}}_{t+1|t} = \sum_i w_i \hat{\chi}_i \ , \tag{8}$$

The variability of the sigma points is calculated as:

$$\mathbf{P}_{\mathbf{xx}} = \sum_i w_i (\hat{\chi}_i - \bar{\mathbf{x}}_{t+1|t})(\hat{\chi}_i - \bar{\mathbf{x}}_{t+1|t})^T + \mathbf{Q}, \tag{9}$$

where $\mathbf{Q}$ is the injected process noise bias used to ensure a non-null spread of sigma points and a positive-definite covariance. This procedure comprises the unscented transform used to estimate the behavior of a nonlinear function.

To obtain the predicted observation, we again apply the unscented transform, this time using the predicted states, $\mathbf{X}_{t+1|t}$, to estimate what we expect to observe from the hypothetical measurement of each state: $\varphi = v[\chi] \in \mathbb{R}^k$. This way we obtain the predicted set of observations, $\mathbf{Y}_{t+1|1} = \{v[\hat{\chi}_i]\} = \{\hat{\varphi}_i\}$, and may calculate its weighted mean and covariance,

$$\begin{aligned} \bar{\mathbf{y}}_{t+1|t} &= \sum_i w_i \hat{\varphi}_i, \\ \mathbf{P}_{\mathbf{yy}} &= \sum_i w_i (\hat{\varphi}_i - \bar{\mathbf{y}}_{t+1|t})(\hat{\varphi}_i - \bar{\mathbf{y}}_{t+1|t})^T + \mathbf{R}, \end{aligned} \tag{10}$$

where $\mathbf{R}$ is the injected measurement noise bias again used to ensure a positive-definite covariance. The cross correlation between the estimated state and estimated measurement may also be calculated:

$$\mathbf{P}_{\mathbf{xy}} = \sum_i w_i (\hat{\chi}_i - \bar{\mathbf{x}}_{t+1|t})(\hat{\varphi}_i - \bar{\mathbf{y}}_{t+1|t})^T. \tag{11}$$

Algorithm 1 Main loop repeated for each fiber

repeat
Form the sigma points $\mathbf{X}_t$ around $\mathbf{x}_t$
Predict the new sigma points $\mathbf{X}_{t+1|t}$ and observations $\mathbf{Y}_{t+1|t}$
Compute weighted means, $\bar{\mathbf{x}}_{t+1|t}$, and $\bar{\mathbf{y}}_{t+1|t}$
Compute covariances, $\mathbf{P}_{\mathbf{xx}}, \mathbf{P}_{\mathbf{xy}}, \mathbf{P}_{\mathbf{yy}}$
Update estimate $(\mathbf{x}_{t+1}, \mathbf{P}_{t+1})$ using scanner measurement $\mathbf{y}_{t+1}$
Proceed in the estimated direction $\mathbf{m}_{t+1}$
until estimated model appears isotropic

As is done in the classic linear Kalman filter, the final step is to use the Kalman gain,

$$\mathbf{K} = \mathbf{P}_{\mathbf{xx}}\mathbf{P}_{\mathbf{yy}}^{-1}, \tag{12}$$

to correct our prediction and provide us with the final estimated system state, and with a covariance estimate to be used in the next steps:

$$\mathbf{x}_{t+1} = \bar{\mathbf{x}}_{t+1|t} + \mathbf{K}(\mathbf{y}_{t+1} - \bar{\mathbf{y}}_{t+1|t}) \tag{13}$$

$$\mathbf{P}_{t+1} = \mathbf{P}_{\mathbf{xx}} - \mathbf{K}\mathbf{P}_{\mathbf{yy}}\mathbf{K}^T . \tag{14}$$

To initialize this process we use $\mathbf{P}_0 = \mathbf{I}^{n\times n}$, and $\mathbf{x}_0$ is initialized by a single tensor estimation out of the acquired signal, y_0.

We continue in this manner until the model appears isotropic. Algorithm 1 outlines these steps.

4.3 Free-Water and Tensor Corrected Maps Following Filtered Tractography

The previous section outlines the estimation of the free-water model using filtered tractography, however, any model could be estimated using this approach (see in [26] for a number of different models that were included in the filtered tractography approach). We generated FA and FA_t maps based on a full brain tractography of a healthy volunteer. The maps were generated by recording the estimated model parameters for each voxel while performing the filtered tractography. Running the free-water model yielded free-water maps and FA_t maps. Running the filtered tractography on a model that does not include the fractional volume parameter yielded the DTI model and hence FA maps. Unlike the maps produced by the single- and multi-shell fit, these maps have values only in places where the tractography method identified fiber tracts. Therefore, the filtered tractography is limited to provide maps of white-matter alone.

Figure 3 compares axial slices of the FA and FA_t along with a difference map. Adding the free water term only adds small qualitative changes to the FA map. Looking at the difference maps (third from left) it can be seen that FA_t is overall

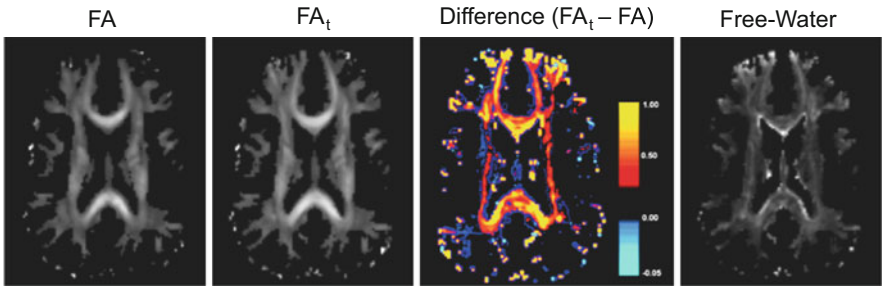


Fig. 3 Filtered tractography free-water maps. Maps of FA and FA$_t$ were generated while tracking white matter tracts in the entire brain. These maps are comparable to the single- and multi-shell maps, however, they have values only in white matter. The contrast of FA ranges between [0,1] and FA$_t$ is in general higher than FA especially in areas with significant partial volume, that have higher free-water values, as can be seen in the difference map and the free-water map

higher than FA. The biggest differences are observed around areas with partial volumes of CSF, around the ventricles and close the cortical areas. These areas match the ones with high free-water contents as can be seen from the free water map in the Fig. 3 (right). We observed, that including the free water term always increases the FA, decreases the Trace, and better fits the signal. The average free water content of all fibers was 19.09 %.

Next we compared the diffusion models on brain scans of 10 healthy human subjects. To generate the fiber tracts we began by seeding each voxel once, and traced the fibers for each of the diffusion models until the anisotropy threshold was reached, to arrive at a full brain tractography. From there we extracted the fibers that pass through the anterior limb of the internal capsule to filter out the respective tract under investigation. Fibers crossing the brain stem were excluded. We chose to perform full-brain tractography first in order to obtain more complete tract reconstructions. The fibers were expected to connect to the frontal gyri (green in Fig. 4). Furthermore, the fibers are expected not to be connected to the adjoining areas of precentral gyrus, and the caudal-middle-frontal gyrus (red in Fig. 4). The resulting tracts for all methods are shown in Fig. 4. Including free-water in the estimation produces a more complete fiber than without including free-water. However, the tracts did not reach all of the frontal gyri. When modifying the model to include three compartments, two with fiber tensors, and one with free-water, a much more complete delineation is achieved. This is suggesting that these fibers are affected by both partial volume with extra-cellular space, and crossing-fibers. The addition of the free-water to the two-tensor model produced slightly more complete fiber than without the free-water component, however, it also introduced a false-positive fiber that connected the cortico-spinal tracts with the caudal-middle-frontal gyrus.

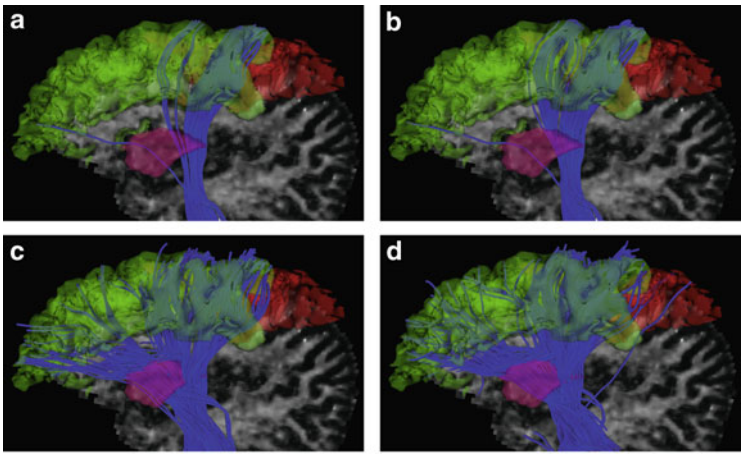


Fig. 4 Comparison of filtered tractography models. Fiber bundles passing through the anterior limb of the internal capsule (*purple*) to the target region (*green*), and the non-target region (*red*), as generated for four different filtered tractography models. Accounting for crossing-fibers by adding a second tensor component produces a much more complete fiber. Nevertheless, adding free-water provides an ever more complete representation of the fibers. (**a**) *DTI*. (**b**) *Free-water*. (**c**) *2-tensor*. (**d**) *2-tensor + free-water*

5 Free Water and Group Comparisons

Controlling for partial volume is especially important when comparing groups of subjects where changes in the volume might be part of the pathology. When performing group comparisons, it is usually advised to add the volume as a covariate, or to find analysis methods that obviate the volume changes all together [27]. One example for such an approach is tract based spatial statistics (TBSS) [28], which projects all the diffusivities onto a skeleton of the white-matter, thus representing the center of the fiber, which is supposed to be free of partial volume effects with the surrounding tissue. While this technique avoids regions of the fiber that are more prone to include partial volume effects and thus improves the accuracy of the statistics, at the same time many of the voxels are ignored, thus limiting the precision and statistical power of the analysis and potentially missing important information that might be in the data.

Alternatively, region-of-interest (ROI) analysis can manually achieve robust and accurate measurements that include as many voxels in the target region as possible. However, the ROI approach is prone to cause high inter- and intra-observer variability. In addition, in order to ensure accuracy and avoid the discussed biases due to partial volume, the regions have to be drawn in a very conservative fashion without the inclusion of surrounding structures or CSF. This again can lead to decreased precision and statistical power of the overall analysis.

In this section we demonstrate how free-water elimination can assist in controlling for partial volume in the delineation of ROIs. We apply a partial volume

clustering approach that combines a histogram based clustering analysis with an atlas-based placement of ROIs to avoid manual ROI placement and to circumvent the decreased sensitivity occurred when limiting the analysis to the white matter skeleton. The application of free-water enhances the partial volume clustering and provides the ability to differentiate macroscopic shrinking effects (i.e., atrophy) from microscopic alterations (such as cellular density changes). As a result we are able to compare microstructural changes that occur not only at the center but also at the periphery of fibers.

5.1 Partial Volume Clustering

Partial volume clustering is a technique for the robust extraction of diffusion indices from fiber bundles. It employs a probabilistic mixture model for differentiating fiber voxels from isotropic background, similarly to the procedure described in [29]. If a region includes both fiber and non fiber components, then using a probabilistic mixture model and histogram analysis, the ROI can be segmented into a fiber, a non fiber and a partial volume class on basis of a scalar anisotropy value that is derived on a voxel basis.

The classification algorithm assumes that the signal in a voxel is composed of signal from a fiber class (F), an isotropic background class (B), and a mixture class (M) [29]. The partial volume that the three classes occupy in the ROI is denoted by π_F, π_B and π_M respectively such that $\pi_F + \pi_B + \pi_M = 1$. The signal of the mixture class is given by

$$S_M = (1-\rho)S_F + \rho S_B,$$

where ρ is the mixture parameter. The anisotropy index $a = 1 - (\lambda_2 + \lambda_3)/(2\lambda_1)$ [29, 30] is used as measure for the anisotropy, with λ_1, λ_2, and λ_3 being the sorted eigenvalues of the diffusion tensor. Let $P(a|F)$, $P(a|B)$ and $P(a|M)$ be the conditional probabilities that a is measured in fiber, background and mixture class. Then, the probability to measure a is

$$P(a) = \pi_F P(a|F) + \pi_B P(a|B) + \pi_M P(a|M)\,.$$

Here, $P(a|F)$ and $P(a|B)$ are modeled by Gaussian distributions with mean values m_F and m_B and variances σ_F and σ_B. By further assuming that ρ is uniformly distributed, $P(a|M)$ can be modelled [31] by

$$\begin{aligned} P(a|M) = &\frac{1}{\sqrt{2\pi}} \int_0^1 \frac{d\rho}{\sqrt{\rho^2\sigma_B^2 + (1-\rho)^2\sigma_F^2}} \\ &\cdot \exp\left(-\frac{((1-\rho)m_F + \rho m_B - a)^2}{2((1-\rho)^2\sigma_F^2 + \rho^2\sigma_B^2)}\right) \end{aligned} \tag{15}$$

For the evaluation of a ROI, the free parameters $m_F, m_B, \sigma_F, \sigma_B, \pi_F, \pi_B$ and π_M must be determined. Thereto, the a-values of all voxels within the ROI are arranged in a histogram. Then, the free parameters are determined using a generalization of the EM algorithm [32] that was introduced by Laidlaw et al. [33]. Now, the conditional probabilities $P(a|F)$, $P(a|B)$ and $P(a|M)$ are known. Using Bayes' rule, we can find the class probabilities, for example for the fiber class,

$$P(F|a) = \pi_F P(a|F)/P(a) \, .$$

Then we can also calculate the expectation value for any arbitrary value V (e.g. for FA values) for a given class across a ROI, Ω. For example, the expectation value within the fiber class in a ROI of size n voxels is:

$$\langle V \rangle = \frac{1}{n\pi_F} \sum_{i \in \Omega} P(F|a_i) V_i$$

6 Partial Volume and Free-Water in the Corpus Callosum of Alzheimer's Disease

As a demonstration of the partial-volume clustering, we consider a problem of group comparison between a clinical population of Alzheimer's disease (AD) patients and normal controls. The disease manifests as a severe form of dementia and its signature pathologies are plaques and tangles. However, this neurodegenerative disease is also causing global alteration to the white-matter integrity via processes of inflammation and demyelination [34]. The Corpus-Callosum is known to be one of the prime white matter fibers to be affected in Alzheimer's [35]. To test the partial volume clustering we check if it is sensitive enough to identify the abnormalities on the Corpus-Callosum, and compare the results with the sensitivity of the TBSS approach.

6.1 Data Acquisition

A single shell (twice refocused) DTI data was collected on a 1.5 T (Symphony, Siemens) from 15 AD patients and 15 matched healthy controls. The data was collected using the following parameters: TR/TE 4,700/78 ms, FOV 240 mm, data matrix of 96×96 yielding an in-plane resolution of 2.5 mm, 50 axial slices with a thickness of 2.5 mm and no gap, with 6 gradient directions ($b = 1{,}000\,\mathrm{s/mm^2}$) and a b_0 image. This scheme was repeated 10 times. All images were corrected for motion and eddy currents (FSL, FLIRT), while compensating the gradient directions. Images were masked (FSL, BET) and the tensor toolkit was used for

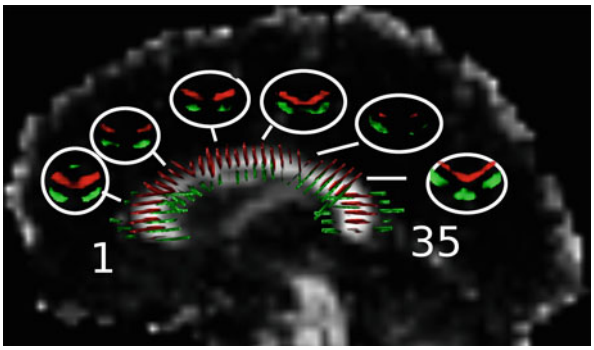


Fig. 5 Partial Volume Clustering. The Corpus Callosum was segmented into 35 ROIs along the mid-sagittal space. Partial volume clustering was run in each ROI, revealing that there was partial volume of the Corpus Callosum with its surrounding tissue. The "fiber" class is colored in *red*, "non-fiber" or background is *green* and "partial-volume" in *black*. Following clustering we use all of the "fiber" class to compare between subjects. Other methods, such as tract based spatial statistics (TBSS) use a subset of the "fiber" class (example, the skeleton) to conduct group comparisons. Including the entire "fiber" class in the analysis increases the number of samples and the statistical power (compare Fig. 6 for TBSS with Fig. 8 for the clustering approach)

tensor estimation (https://gforge.inria.fr/projects/ttk). Free-water corrected tensors and free-water maps were calculated using the single-shell estimation approach. From the tensors maps we extracted the radial diffusivity, $(\lambda_2 + \lambda_3)/2$, resulting with radial diffusivity maps and free-water corrected radial diffusivity maps that we used in order to compare between the groups.

6.2 Partial Volume Clustering Versus TBSS

Radial diffusivity is expected to be sensitive to pathological processes such as demyelination, that are common in AD, and we therefore compared the sensitivity of our partial-volume clustering approach to radial diffusivity abnormalities with the sensitivity of a TBSS approach.

For the clustering analysis, registration was performed in three steps directly on the tensor datasets using DTITK (http://www.nitrc.org/projects/dtitk):

1. A template was bootstrapped using the IXI aging template.
2. A population specific template space was created using affine and diffeomorphic registration.
3. Diffusion indices were extracted from the subjects by employing the probabilistic mixture model that was described in Sect. 5.1.

Thirty five fairly large ROIs were manually defined in template space using the open-source MITK Diffusion 2011 (www.mitk.org), and then placed along the Corpus Callosum using atlas based positioning, as illustrated in Fig. 5. The ROIs

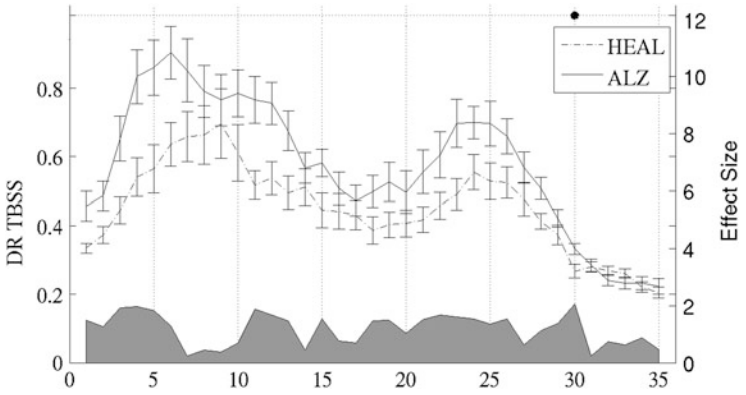


Fig. 6 Non-corrected radial diffusivities. Group differences of radial diffusivities (DR) between AD patients and normal controls based on TBSS analysis, show a single ROI that is significantly ($p < 0.05$) abnormal in AD (indicated with a *black circle*)

were defined in such a way that they parcellate the Corpus Callosum to even sections along the Corpus Callosum contour as it is found in the mid-sagittal slice. These ROIs included the Mid-Sagittal portion of the Corpus Callosum along with the tissue that surrounds it, thus each ROI included a partial volume of fiber, and non-fiber classes.

In addition to the partial volume clustering, the full TBSS pipeline was applied using the parameters suggested in [28]. The projection of the TBSS skeleton was applied to obtain radial diffusivities and free-water eliminated radial diffusivities skeletons. The Corpus-Callosum was identified in the mid-sagittal slice, and was linearly divided into 35 equal segments, to match the ROIs defined for the clustering approach.

We compared the two groups (AD and controls) separately for each ROI, using a two-sample unpaired t-test controlled for age with $p = 0.05$ as the threshold for significance.

6.3 Sensitivity to Abnormalities in Alzheimer's Disease

We first investigate abnormalities between the AD patients and the normal controls as they appear on the skeleton of the Corpus Callosum (using the TBSS analysis). We compare the sensitivity of the regular DTI radial diffusivity measure (Fig. 6), and compare it with the free-water corrected radial diffusivity (Fig. 7). We find that in general the radial diffusivity (either corrected or not corrected for free-water) in AD is higher than in controls. However, this difference was significant (indicated by black circles) in only one location when using non-corrected values, and in four locations when using corrected values. This is suggesting that in general the

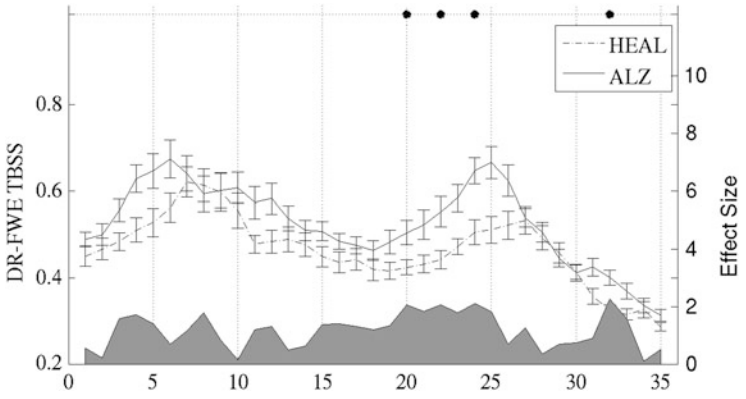


Fig. 7 Free-Water corrected radial diffusivities. Group differences of radial diffusivities corrected for free-water (DR-FWE) based on TBSS analysis reveal four ROIs with significant abnormalities. All the ROIs are in the posterior part of the Corpus-Callosum

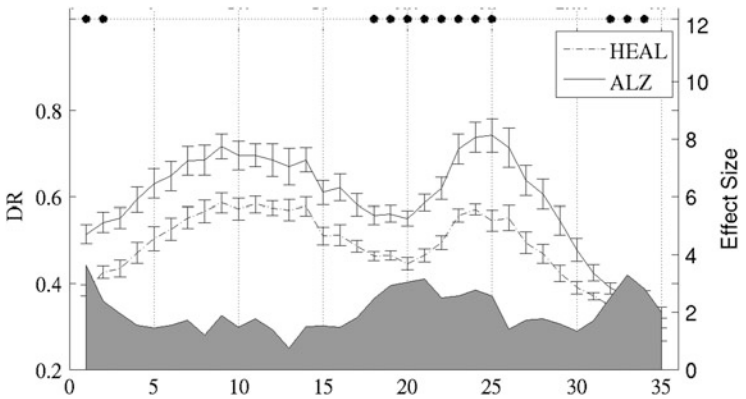


Fig. 8 Partial volume clustering of non-corrected radial diffusivity. Looking at the "fiber" class reveals many significant abnormalities ($p < 0.05$) forming two continuous clusters in the posterior part of the Corpus-Callosum. Looking on the "fiber" class reveals many more abnormalities than looking at the skeleton (see Fig. 6)

TBSS analysis is not extremely sensitive to the underlying pathology of AD, yet the sensitivity is increased when correcting for free-water. This is probably due to the fact that the corrected values are smoother than the non-corrected values.

We next compare the differences between AD patients and normal controls over the entire fiber cluster (as found by the partial volume clustering technique). We compare the non-corrected radial diffusivity values (Fig. 8) with the free-water corrected radial diffusivity values (Fig. 9). We find that the general trend of higher radial diffusivity in AD is preserved, however now there are many more ROIs in the Corpus-Callosum of AD that have a statistically significant increase. These ROIs are all located in the posterior region of the Corpus-Callosum (ROIs 18-25), suggesting

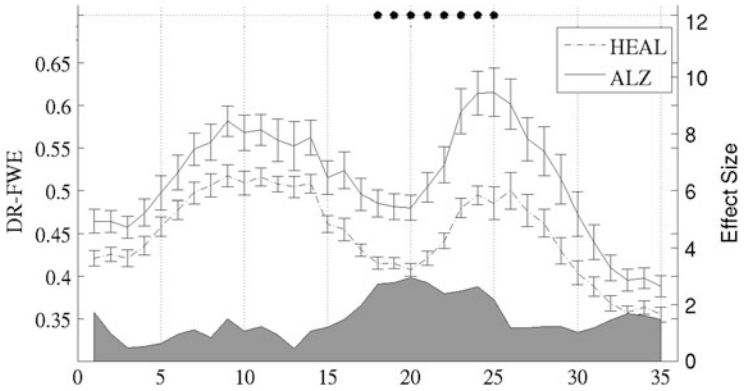


Fig. 9 Partial volume clustering of free-water corrected radial diffusivity. Consistent with the non-corrected findings (Fig. 6), many abnormalities form a cluster in the posterior part of the Corpus-Callosum, suggesting a demyelination related abnormality. The second cluster does not appear in the free-water analysis, suggesting that the underlying pathology there could be neuroinflammation or atrophy

that this part of the fiber has the most pathology. The increased sensitivity comparing to TBSS suggests that the pathology is mainly in the perimeters of the fiber, and not at its center, which is expected in cases of atrophy. Finally, the abnormal region can be divided to two parts, one that consistently shows abnormalities in both corrected and non-corrected values, and a second that is abnormal only in the non-corrected radial diffusivity values. This suggests the possibility of two different underlying pathology, demyelination, that increases radial diffusivity in both corrected and non-corrected values, versus inflammation, which is expected to increase radial diffusivity, but not corrected radial diffusivity.

7 Summary

Fitting the free-water model instead of the DTI model adds the extracellular volume as a new estimated biological parameter and provides tensor images that are more tissue specific [8–11]. This comes with a computational price, as the model is harder to estimate. The three approaches (single-shell, multi-shell, filtered tractography) discussed here provide means for estimation of the model parameters. This chapter covers the technical details involved in the estimation of the free-water maps and free-water corrected tensors, however, a systematic comparison is still in need, in order to better define when each method should be preferred.

It seems that the decision of which method to use should be driven by the type of data available, and the research question in mind. Specifically, if the research design can afford extra scan time, or if a multi-shell data is available, using the multi-shell estimation will be more stable and potentially more accurate then the single-shell

approach. With a multi-shell data we can relax assumptions and dependency on regularization that is required when fitting the model from a single-shell data. If the research question is specific to a certain white-matter bundle, then using the filtered tractography is a better selection. Of note, even though we presented filtered-tractography results on single-shell data, the method is straightforward to generalize for the multi-shell data.

Future work should also test the accuracy of the estimated parameters, through specially designed phantoms, or animal model scans. For example, the free-water estimation is biased by differences in T1 and T2 [9], it could be affected by temperature changes, and it would be interesting to see how it is affected by the types of noise that is expected in the dMRI acquisition. Another important task is to demonstrate pathological correlates of the free-water parameters. For example, it is clear that neuroinflammation and demyelination affects changes in the free-water and FA_t respectively, yet the specificity of these measures to the pathologies is not yet clear.

Since the introduction of the free-water model there were a number of higher order models that included a compartment of free-water with fixed diffusivities, along with various other compartments. For example, the The composite hindered and restricted water diffusion (CHARMED) model [36], the Axcaliber model [37], and more recently the neurite orientation dispersion and density imaging (NODDI) model [38]. Since it is easier to relate the free-water compartment to a biological compartment, the inclusion of this compartment in these models increase the specificity of the remaining compartments to the underlying tissue. These higher order methods typically require specialized acquisition, and it is yet to be determined whether the free-water compartment estimated in these measures matches the estimation of the free-water model as described here, although some preliminary results suggest that the values agree [39].

It is nevertheless clear that similar to the higher order models, the inclusion of the free-water parameter in the diffusion tensor model adds valuable new information to the study of the nervous system. We have demonstrated here the extra information that can be obtained in tractography, and in studying abnormalities related to AD, which is similarly applicable to many other brain disorders as well as for the study of normal development and aging.

Acknowledgements This work was supported by the following grants: Department of Defense X81XWH-07-CC-CSDoD; NIH P41RR013218, P41EB015902, NIH R01MH074794; VA Merit Award. OP is partly supported by a National Alliance for Research on Schizophrenia and Depression (NARSAD) Young Investigator Grant from the Brain and Behavior Research Foundation.

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**, 259–267 (1994)
2. Assaf, Y., Pasternak, O.: Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. J. Mol. Neurosci. **34**(1), 51–61 (2008). doi:10.1007/s12031-007-0029-0

3. Alexander, A.L., Lee, J.E., Lazar, M., Field, A.S.: Diffusion Tensor Imaging of the Brain. Neurotherapeutics **4**(3), 316–329 (2007). doi:10.1016/j.nurt.2007.05.011
4. Vos, S.B., Jones, D.K., Viergever, M.A., Leemans, A.: Partial volume effect as a hidden covariate in DTI analyses. NeuroImage **55**(4), 1566–1576 (2011). doi:10.1016/j.neuroimage.2011.01.048
5. Pasternak, O., Assaf, Y., Intrator, N., Sochen, N.: Variational multiple-tensor fitting of fiber-ambiguous diffusion-weighted magnetic resonance imaging voxels. Magn. Reson. Imaging **26**(8), 1133–1144 (2008)
6. Malcolm, J.G., Shenton, M.E., Rathi, Y.: Filtered multi-tensor tractography. IEEE Trans. Med. Imaging **29**, 1664–1675 (2010). doi:10.1109/TMI.2010.2048121
7. Alexander, D.: Multiple-fibre reconstruction algorithms for diffusion MRI. Ann. N. Y. Acad. Sci. **1046**, 113–133 (2005)
8. Jones, D.K., Cercignani, M.: Twenty-five pitfalls in the analysis of diffusion MRI data. NMR Biomed. **23**(7), 803–820 (2010). doi:10.1002/nbm.1543
9. Pasternak, O., Sochen, N., Gur, Y., Intrator, N., Assaf, Y.: Free water elimination and mapping from diffusion mri. Magn. Reson. Med. **62**(3), 717–730 (2009)
10. Metzler-Baddeley, C., O'Sullivan, M.J., Bells, S., Pasternak, O., Jones, D.K.: How and how not to correct for CSF-contamination in diffusion MRI. NeuroImage **59**(2), 1394–1403 (2012). doi:10.1016/j.neuroimage.2011.08.043
11. Wang, Y., Wang, Q., Haldar, J.P., Yeh, F.C., Xie, M., Sun, P., Tu, T.W., Trinkaus, K., Klein, R.S., Cross, A.H., Song, S.K.: Quantification of increased cellularity during inflammatory demyelination. Brain **134**(12), 3590–3601 (2011). doi:10.1093/brain/awr307
12. Metzler-Baddeley, C., Jones, D., Belaroussi, B., Aggleton, J., O'Sullivan, M.: Frontotemporal connections in episodic memory and aging: A diffusion MRI tractography study. J. Neurosci. **31**(37), 13236–13245 (2011)
13. Pasternak, O., Westin, C.F., Bouix, S., Seidman, L.J., Goldstein, J.M., Woo, T.U.W., Petryshen, T.L., Mesholam-Gately, R.I., McCarley, R.W., Kikinis, R., et al.: Excessive extracellular volume reveals a neurodegenerative pattern in schizophrenia onset. J. Neurosci. **32**(48), 17365–17372 (2012)
14. Fritzsche, K., Stieltjes, B., van Bruggen, T., Meinzer, H.P., Westin, C.F., Pasternak, O.: A combined approach for the elimination of partial volume effects in diffusion MRI. In: Proceedings of the 20th ISMRM, Melbourne, p. 3548 (2012)
15. Harris, K.R., Woolf, L.A.: Pressure and temperature dependence of the self diffusion coefficient of water and oxygen-18 water. J. Chem. Soc. Faraday Trans. 1 **76**, 377–385 (1980). doi:10.1039/F19807600377, http://dx.doi.org/10.1039/F19807600377
16. Pierpaoli, C., Jones, D.: Removing CSF contamination in brain DT-MRIs by using a two-compartment tensor model. In: Proceedings of the 12th ISMRM, Kyoto, p. 1215 (2004)
17. Mulkern, R.V., Haker, S.J., Maier, S.E.: On high b diffusion imaging in the human brain: ruminations and experimental insights. Magn. Reson. Imaging **27**(8), 1151–1162 (2009). doi:10.1016/j.mri.2009.05.003
18. Behrens, T.E., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M.: Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. **50**(5), 1077–1088 (2003)
19. Pierpaoli, C., Jezzard, P., Basser, P., Barnett, A., Di Chiro, G.: Diffusion tensor MR imaging of the human brain. Radiology **201**(3), 637–648 (1996)
20. Gur, Y., Pasternak, O., Sochen, N.: Fast GL(n)-invariant framework for tensors regularization. Int. J. Comput. Vis. **85**(3), 211–222 (2009)
21. Pasternak, O., Sochen, N., Basser, P.J.: The effect of metric selection on the analysis of diffusion tensor MRI data. NeuroImage **49**(3), 2190–2204 (2010). doi:10.1016/j.neuroimage.2009.10.071, http://dx.doi.org/10.1016/j.neuroimage.2009.10.071
22. Jones, D.K.: The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a monte carlo study. Magn. Reson. Med. **51**(4), 807–815 (2004)

23. Pasternak, O., Shenton, M., Westin, C.F.: Estimation of extracellular volume from regularized multi-shell diffusion MRI. In: Proceedings of the MICCAI, Nice, pp. 305–312 (2012)
24. Westin, C.F., Pasternak, O., Knutsson, H.: Rotationally invariant gradient schemes for diffusion MRI. In: Proceedings of the 20th ISMRM, Melbourne (2012)
25. King, M.D., Gadian, D.G., Clark, C.A.: A random effects modelling approach to the crossing-fibre problem in tractography. NeuroImage **44**, 753–768 (2009)
26. Baumgartner, C., Michailovich, O., Levitt, J., Pasternak, O., Bouix, S., Westin, C., Rathi, Y.: A unified tractography framework for comparing diffusion models on clinical scans. In: Computational Diffusion MRI Workshop of MICCAI, Nice, pp. 27–32 (2012)
27. Metzler-Baddeley, C., O'Sullivan, M.J., Bells, S., Pasternak, O., Jones, D.K.: How and how not to correct for CSF-contamination in diffusion MRI. NeuroImage **59**(2), 1394–1403 (2012). doi:10.1016/j.neuroimage.2011.08.043
28. Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.J.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage **31**(4), 1487–1505 (2006). doi:10.1016/j.neuroimage.2006.02.024, http://dx.doi.org/10.1016/j.neuroimage.2006.02.024
29. Schlueter, M., Stieltjes, B., Hahn, H.K., Rexilius, J., Konrad-verse, O., Peitgen, H.O.: Detection of tumour infiltration in axonal fibre bundles using diffusion tensor imaging. Int. J. Med. Robot. **1**(3), 80–86 (2005). doi:10.1002/rcs.31, http://dx.doi.org/10.1002/rcs.31
30. Stieltjes, B., Schlüter, M., Didinger, B., Weber, M.A., Hahn, H.K., Parzer, P., Rexilius, J., Konrad-Verse, O., Peitgen, H.O., Essig, M.: Diffusion tensor imaging in primary brain tumors: reproducible quantitative analysis of corpus callosum infiltration and contralateral involvement using a probabilistic mixture model. NeuroImage **31**(2), 531–542 (2006). doi:10.1016/j.neuroimage.2005.12.052, http://dx.doi.org/10.1016/j.neuroimage.2005.12.052
31. Noe, A., Gee, J.C.: Partial volume segmentation of cerebral mri scans with mixture model clustering. In: IPMI, Davis, pp. 423–430 (2001)
32. McLachlan, G.J., Krishnan, T.: The EM Algorithm and Extensions. Wiley Series in Probability and Statistics (2007). doi:10.1002/9780470191613
33. Laidlaw, D.H., Fleischer, K.W., Barr, A.H.: Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms. IEEE Trans. Med. Imaging **17**(1), 74–86 (1998). doi:10.1109/42.668696, http://dx.doi.org/10.1109/42.668696
34. Weiner, H.L., Selkoe, D.J.: Inflammation and therapeutic vaccination in CNS diseases. Nature **420**(6917), 879–884 (2002). doi:10.1038/nature01325, http://www.nature.com/nature/journal/v420/n6917/full/nature01325.html
35. Agosta, F., Pievani, M., Sala, S., Geroldi, C., Galluzzi, S., Frisoni, G.B., Filippi, M.: White matter damage in alzheimer disease and its relationship to gray matter atrophy. Radiology **258**(3), 853–863 (2011). doi:10.1148/radiol.10101284, http://radiology.rsna.org/content/258/3/853.long
36. Assaf, Y., Basser, P.J.: Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. NeuroImage **27**(1), 48–58 (2005). doi:10.1016/j.neuroimage.2005.03.042, http://www.sciencedirect.com/science/article/pii/S1053811905002259
37. Assaf, Y., Blumenfeld-Katzir, T., Yovel, Y., Basser, P.J.: Axcaliber: a method for measuring axon diameter distribution from diffusion MRI. Magn. Reson. Med. **59**(6), 1347–1354 (2008)
38. Zhang, H., Schneider, T., Wheeler-Kingshott, C.A.M., Alexander, D.C.: NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. NeuroImage **61**(4), 1000–1016 (2012)
39. van Bruggen, T., Zhang, H., Pasternak, O., Meinzer, H.P., Stieltjes, B., Fritzsche, K.H.: Free-water elimination for assessing microstructural gray matter pathology - with application to alzheimer's disease. In: Proceedings of the 21th ISMRM, Salt Lake City, p. 790 (2013)

Techniques for Computing Fabric Tensors: A Review

Rodrigo Moreno, Magnus Borga, and Örjan Smedby

Abstract The aim of this chapter is to review different approaches that have been proposed to compute fabric tensors with emphasis on trabecular bone research. Fabric tensors aim at modeling through tensors both anisotropy and orientation of a material with respect to another one. Fabric tensors are widely used in fields such as trabecular bone research, mechanics of materials and geology. These tensors can be seen as semi-global measurements since they are computed in relatively large neighborhoods, which are assumed quasi-homogeneous. Many methods have been proposed to compute fabric tensors. We propose to classify fabric tensors into two categories: mechanics-based and morphology-based. The former computes fabric tensors from mechanical simulations, while the latter computes them by analyzing the morphology of the materials. In addition to pointing out advantages and drawbacks for each method, current trends and challenges in this field are also summarized.

1 Introduction

One of the ultimate goals of trabecular bone research in medicine is to determine the effect of pathological conditions of trabecular bone, such as osteoporosis and osteoarthritis, and their treatments on the quality of trabecular bone. One of the parameters that can be used to evaluate the bone quality is its anisotropy.

R. Moreno (✉) • Ö. Smedby
Department of Medical and Health Sciences (IMH), Center for Medical Image Science and Visualization (CMIV), Linköping University, Campus US, 58185 Linköping, Sweden
e-mail: rodrigo.moreno@liu.se; orjan.smedby@liu.se

M. Borga
Department of Biomedical Engineering (IMT), Center for Medical Image Science and Visualization (CMIV), Linköping University, Campus US, 58185 Linköping, Sweden
e-mail: magnus.borga@liu.se

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_12,

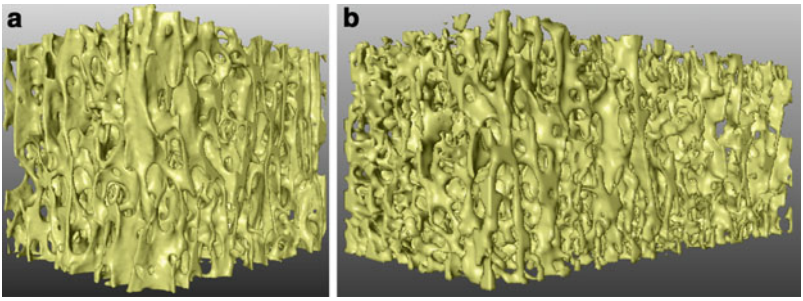


Fig. 1 Rendering of scans of trabecular bone from a radius and a vertebra respectively acquired through micro computed tomography

For example, evidences supporting that changes in the anisotropy and orientation of trabecular bone are associated with osteoporosis has been reported [11, 32, 47]. Figure 1 show renderings from two in vitro specimens.[1]

Trabecular bone is a tissue that is under continuous remodeling [57, 67, 68]. This remodeling process, which is driven by both physiology and mechanical adaptation processes, usually generates anisotropies in trabecular bone. Since mechanical stimuli differs from site to site of the body, analyses of changes over time of anisotropy generated by non-mechanical causes are performed site-dependent. In this context, fabric tensors are a fundamental tool to perform such kind of analyses.

Fabric tensors aim at modeling through tensors both anisotropy and orientation of a material of interest (usually referred to as *phase* in mechanics of materials) with respect to another one. In trabecular bone research, these two phases correspond to trabecular bone and bone marrow respectively. In addition to trabecular bone, fabric tensors have been used in other fields, such as mechanics of materials [91] and geology [49]. Fabric tensors are semi-global measurements in the sense that they are computed in relatively large neighborhoods, which are assumed quasi-homogeneous. In mechanics, such neighborhoods are usually referred to as representative volume elements (RVE) [72]. Furthermore, since it has been shown that microstructural architecture of most materials, including trabecular bone, can be accurately modeled by means of second-order tensors [48, 94], higher-order tensors are usually not computed.

In this context, the aim of this chapter is to review different approaches that have been proposed for computing fabric tensors, pointing out their advantages and disadvantages. We propose to classify these approaches into two categories: mechanics-based and morphology-based. The former approach computes fabric

[1]We thank Prof. Osman Ratib from the Service of Nuclear Medicine at the Geneva University Hospitals for providing the μCT scan of the vertebra; Andres Laib from SCANCO Medical AG and Torkel Brismar from the Division of Radiology at the Karolinska University Hospital for providing the μCT scan of the radius.

tensors from mechanical simulations, while the latter computes them by analyzing the morphology of trabecular bone. It is important to remark that, although some authors do not consider tensors computed through mechanical simulations as a specific type of fabric tensors, we argue they actually are fabric tensors, since they can also be used to describe orientation and anisotropy of trabecular bone, which is the main purpose of fabric tensors. Invariably, the input of all methods is an RVE and the output is the fabric tensor associated to it.

The chapter is organized as follows. The next two sections review the most important methods that follow the aforementioned approaches. Section 4 reviews the research to relate fabric tensors computed through morphology analyses and mechanical properties of the bone. Finally, Sect. 5 makes some concluding remarks, focusing on the current topics in fabric tensors research. As a convention, scalars, vectors and tensors are written in italic, bold and straight font respectively in the paper, e.g. λ, $\mathbf{p}$ and VO.

2 Mechanics-Based Methods

The most relevant property of trabecular bone is its mechanical competence, that is, its capability to bear different types of mechanical loads in different orientations. In this line, mechanics-based methods directly measure fabric tensors from mechanical properties. Since it is difficult to conduct reliable mechanical experiments, these methods compute the tensors through numerical simulations. The next subsections summarize some approaches that follow this path.

2.1 Solid Mechanics Approach

This approach makes use of solid mechanics. A common simplification is to assume that trabecular bone is an elastic material [73]. Thus, under linearity conditions, the so-called stiffness (or elasticity) tensor can directly be used as a fabric tensor. By using the Einstein summation convention, which means that repeated indices imply a summation over them, the stiffness tensor c, can be written as:

$$\boldsymbol{\sigma}_{ij} = \mathsf{c}_{ijk\ell}\, \boldsymbol{\epsilon}_{k\ell}, \tag{1}$$

where σ and ϵ are the stress and strain tensors respectively. Notice that $\boldsymbol{\sigma}$ and $\boldsymbol{\epsilon}$ are of second-order, while c is of fourth-order. This equation corresponds to the generalization of the Hooke's law. Thanks to the symmetries of $\boldsymbol{\sigma}$ and $\boldsymbol{\epsilon}$, c only has 21 out of 81 independent terms. Assuming orthotropic symmetry of trabecular bone [100] the number of independent can be further reduced to nine. By using the Voigt notation [29], c can be represented by the following 6×6 symmetric second-order tensor:

$$\hat{\mathsf{c}} = \begin{pmatrix} \mathsf{c}_{1111} & \mathsf{c}_{1122} & \mathsf{c}_{1133} & \sqrt{2}\mathsf{c}_{1123} & \sqrt{2}\mathsf{c}_{1113} & \sqrt{2}\mathsf{c}_{1112} \\ \mathsf{c}_{1122} & \mathsf{c}_{2222} & \mathsf{c}_{2233} & \sqrt{2}\mathsf{c}_{2223} & \sqrt{2}\mathsf{c}_{2213} & \sqrt{2}\mathsf{c}_{2212} \\ \mathsf{c}_{1133} & \mathsf{c}_{2233} & \mathsf{c}_{3333} & \sqrt{2}\mathsf{c}_{3323} & \sqrt{2}\mathsf{c}_{3313} & \sqrt{2}\mathsf{c}_{3312} \\ \sqrt{2}\mathsf{c}_{1123} & \sqrt{2}\mathsf{c}_{2223} & \sqrt{2}\mathsf{c}_{3323} & 2\mathsf{c}_{2323} & 2\mathsf{c}_{2313} & 2\mathsf{c}_{2312} \\ \sqrt{2}\mathsf{c}_{1113} & \sqrt{2}\mathsf{c}_{2213} & \sqrt{2}\mathsf{c}_{3313} & 2\mathsf{c}_{2313} & 2\mathsf{c}_{1313} & 2\mathsf{c}_{1312} \\ \sqrt{2}\mathsf{c}_{1112} & \sqrt{2}\mathsf{c}_{2212} & \sqrt{2}\mathsf{c}_{3312} & 2\mathsf{c}_{2312} & 2\mathsf{c}_{1312} & 2\mathsf{c}_{1212} \end{pmatrix} \tag{2}$$

The entries of stiffness tensors computed at a local scale can be estimated by running several finite element method (FEM) simulations, at least six [80], each of them with a different boundary condition. Once the local stiffness tensors have been computed, a homogenization procedure can be applied in order to obtain from local estimations a single effective stiffness tensor representative for the whole representative volume element. It has been shown that component-wise addition is not a valid strategy to perform such a homogenization. Thus, more advanced homogenization techniques are usually applied. For example, local structure tensors computed from the relation between local and global strains can be used to steer the homogenization process [9, 30, 31, 80]. Alternatively, Riemannian metrics and the Kullback-Leibler divergence can be applied to aggregate the local tensors [61].

Computing the stiffness tensor through FEM simulations is still under active research [34, 73, 78]. One of the most important problems faced by researchers is that the results can have a large variation for the same sample by applying different boundary conditions, homogenization schemes and methods to generate nodes for the FEM simulations. In addition, another source of error is that the computations are based on the aforementioned simplifications that can be inaccurate. For example, it is well-known that trabecular bone is much better resisting compression than tension [16], while the computed stiffness tensor will predict the same behavior under both boundary conditions. Moreover, most methods are restricted to images acquired from in vitro specimens, as images acquired in vivo have very low quality, which difficults the segmentation required by FEM simulations [40].

2.2 *Wave Propagation Approach*

A more recent approach use FEM simulations of wave propagation on trabecular bone to describe orientation and anisotropy of trabecular bone. Assuming a poroelastic behavior, it has been shown that wave propagation on trabecular bone can be characterized through the acoustic tensor, Q, the solid-fluid interaction tensor, C, and the intrinsic permeability tensor, K, which describe the elastic and viscous effects on the media [14, 15]. Q and C are second-order tensors that are related to the Biot's parameters that describe the stress-strain relation in porous media [12] and the exciting waves. In turn, K, a second-order tensor derived from Darcy's law, takes into account dissipation due to viscous losses and it is closely related to the

tortuosity tensor [76]. A close relation between these tensors and morphology-based fabric tensors has been reported [14].

A shortcoming of this approach is that simulated acoustical properties of trabecular bone have a non-linear dependence on the composition of bone marrow and bone volume fraction, as well as on the resolution of the images [2]. As well as solid mechanics approaches, wave propagation simulations are restricted to high resolution images acquired from in vitro specimens, making it difficult its use in clinical practice.

3 Morphology-Based Methods

Methods that follow this approach compute fabric tensors from the morphology of trabecular bone. These methods have two advantages compared to the methods described in Sect. 2. First, they are largely less computationally expensive than those obtained from mechanical simulations. Second, unlike methods based on mechanics, the resulting fabric tensors are not dependent on the boundary conditions applied during the simulations, homogenization schemes and/or general design of the simulations. However, as a counterpart, it is necessary to relate these fabric tensors with mechanical properties of the material, especially, elasticity.

The vast majority of morphology-based methods use specific features to estimate orientation distributions which are approximated through fabric tensors. If the estimation is performed locally, a homogenization process is applied (usually tensorial summation) in order to obtain global measurements of orientation and anisotropy. The next subsections describe the most important families of methods that follow this approach, which are summarized in Table 1.

3.1 Boundary-Based Methods

Boundary-based methods use the interface between phases to estimate fabric tensors. The Mean Intercept Length tensor (MIL) [83, 94] and the global gradient structure tensor (GST) [6, 88] belong to this category.

3.1.1 Mean Intercept Length Tensor

In trabecular bone research, the MIL tensor is considered as the gold standard thanks to the large amount of evidence supporting its appropriateness to predict mechanical properties of trabecular bone [16, 60, 70, 103]. The MIL tensor was originally proposed as a sampling procedure taken from stereology [83, 94]. The MIL with respect to a particular orientation is defined as the mean distance between a change from one phase to the other in such an orientation. This value

Table 1 Summary of morphology-based methods

Approach	Method
Boundary-based	Mean Intercept Length (MIL) tensor [64, 83, 94]
	Generalized MIL tensor [64]
	Global gradient structure tensor (GST)[6, 59, 88]
Volume-based	Volume orientation tensor (VO) [70]
	Star volume distribution (SVD) [70]
	Star length distribution (SLD) [70]
	Tensor scale [52, 82, 99]
	Inertia tensor [35, 90]
	Sampling sphere orientation distribution (SSOD) [89]
Texture-based	Fractal dimension (FD) [5, 26, 55, 59, 102]
	Hurst orientation transform [77]
	Variance orientation transform [97, 98]
	Line fraction deviation [20, 21]
	Spatial autocorrelation [92]
	Different statistics [27]
Alternative methods	Minkowski tensors [84, 85]
	Diffusion tensor imaging (DTI) [8, 81]
	Texture tensor [23]
	Skeleton-driven [41]
	Assessment of the power spectrum [7]

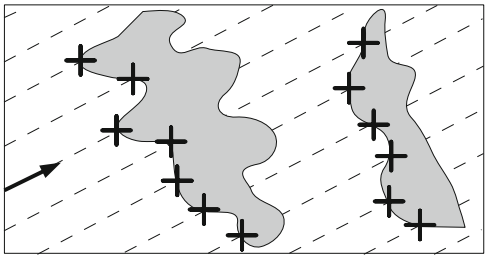


Fig. 2 Computation of the intercepts between a set of parallel lines and the interface between phases. In this example, the number of intercepts is 13

is inversely proportional to the number of intercepts between a set of parallel lines and the interface between phases (see Fig. 2). The MIL tensor is obtained either by applying ellipse/ellipsoid fitting algorithms to polar plots of the MIL computed in different orientations, also known as rose diagrams, or by computing a covariance matrix [38, 64, 87]. Although the orientation distribution of the MIL can also be approximated through higher-order fabric tensors [38], microstructural architecture of most materials can be accurately modeled by means of second-order tensors [48, 94].

Recently, we have proposed a closed formulation for computing the MIL tensor [64]. We have shown that the orientation distribution of intercepts, C_{MIL}, is proportional to the angular convolution between the mirrored extended Gaussian image [33], G, of the sample and the half-cosine kernel, K, which is given by:

$$K(\phi) = \begin{cases} \cos(\phi) \; if \phi \leq \pi/2 \\ 0 \quad otherwise. \end{cases} \tag{3}$$

Thus, C_{MIL} can be computed through:

$$C_{MIL} = \alpha \, G * K, \tag{4}$$

where α is a constant and "$*$" is the angular convolution. Finally, the MIL tensor can be computed by computing a covariance matrix on $1/C_{MIL}$.

This formulation solves several problems of sampling procedures. First, since sampling is avoided, the accuracy is not any longer dependent on the computational cost of the implementation. Second, the new method is not exposed to discretization artifacts generated by line-drawing algorithms. Third, the new formulation is inexpensive since, thanks to the Funk-Hecke theorem [25], the angular convolution can efficiently be computed in the spherical harmonics domain. Fourth, robust implementations of the MIL tensor can straightforwardly be obtained from robust estimations of the extended Gaussian image. Fifth, the new formulation makes straightforward the extension of the MIL tensor to non-binarized images. Finally, the MIL tensor can be generalized by changing different convolution kernels, e.g. to powers of the half-cosine function or the von Mises-Fisher kernel [37].

3.1.2 Global Gradient Structure Tensor

Another boundary-based fabric tensor is the GST [6, 59, 88]. For an image I, the GST is computed as:

$$\mathsf{GST} = \int_{\mathbf{p} \in I} \nabla \mathbf{I_p} \nabla \mathbf{I_p}^T \, dI. \tag{5}$$

Notice that the GST is related to the traditional local structure tensor (ST) [19] computed with a Gaussian of zero mean and standard deviation ρ, K_ρ. If the size of the image is much larger than ρ, the GST can be written as:

$$\mathsf{GST} = \int_{\mathbf{p} \in I} \mathsf{ST} \, dI = \int_{\mathbf{p} \in I} (K_\rho * \nabla \mathbf{I_p} \nabla \mathbf{I_p}^T) \, dI. \tag{6}$$

This method has two interesting properties. First, implementations of the GST are efficient, both in the spatial and frequency domains, and easy to code. Second, the GST and the MIL share the same eigenvectors for binary images [64]. Basically, the GST can be computed as the covariance matrix of $C_{GST} = G * \delta$, where δ is the unit impulse function and G is the mirrored extended Gaussian image. Thus, the difference between the MIL tensor and the GST is that they use a different convolution kernel for computing functions C_{MIL} and C_{GST} and the former calculates the covariance matrix on $1/C_{MIL}$ while the latter on C_{GST}. Hence, both tensors will

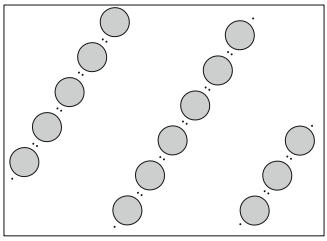


Fig. 3 Example of a material in which boundary-based tensors are unable to estimate anisotropy and orientation. Both, the MIL and GST tensors are isotropic in this case

share the eigenvectors, since both the unit impulse function and the half-cosine kernel are positive and symmetric and changing C by its inverse in the computations does not introduce rotations in the eigenvectors [25].

As an alternative, global structure tensors can also be estimated from local structure tensors computed through quadrature filters [24, 44, 45], by using higher-order derivatives [18, 46] or by means of tensor voting [66].

A drawback of this technique is that the eigenvalues are different to those from the MIL tensor, and the larger one is perpendicular to the main orientation of trabecular bone [88, 96]. Consequently, anisotropies computed through the GST are expected to be in less agreement with the anisotropies yielded by the stiffness tensor [65]. This means that, in practice, the resulting tensor has to be post-processed in order to be used as a predictor of mechanical properties.

3.2 Volume-Based Methods

A problem of the boundary-based methods is that they are only appropriate where the anisotropy and orientation are determined by the interface between phases. For example, boundary-based methods are unable to estimate anisotropy in the case of Fig. 3.

To solve this issue, volume-based methods compute anisotropy from measures taken inside one of the phases. The next subsections describe the most important families of methods that follow this approach.

3.2.1 Distributions of Intercepting Lines

There are many fabric tensors that are computed through the sampling procedure shown in Fig. 4. First, a set of N sampling points are generated in the material of interest (e.g., trabecular bone). Second, the intercept length of lines with different orientations that cross every testing point is computed.

Several features can be extracted in order to compute fabric tensors. For example, local volume orientation at a point is given by the orientation corresponding to the largest intercept at that point. The volume orientation tensor VO, is computed as [70]:

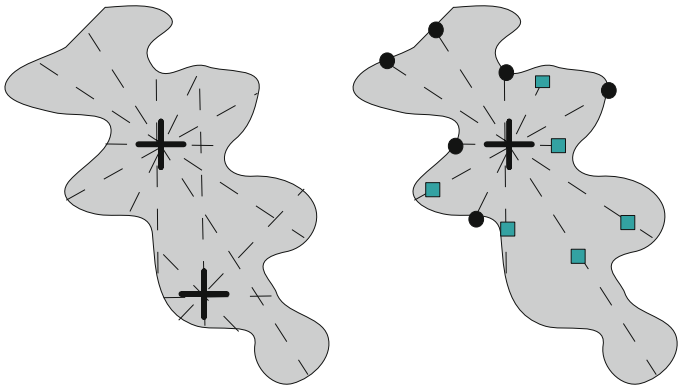


Fig. 4 Distributions of intercepting lines. *Left*: lines with different orientations are traced from some sampling points (marked with *crosses*). The lenght of those lines are used to generate the VO, SVD and SLD tensors. *Right*: in order to compute the scale tensor, line segments are shortened (half of the intercepts with the boundary are shifted to the positions marked with *squares*) in order to make them symmetric with respect to the sampling point

$$\mathsf{VO} = \sum_{i=1}^{N} L_{max_i} \ \mathbf{v}_{max_i} \, \mathbf{v}_{max_i}^T , \tag{7}$$

where L_{max_i} and $\mathbf{v}_{max_i}$ are the largest intercept at i and its corresponding orientation respectively. On the other hand, the star volume (SVD) and length distributions (SLD) consider all intercepts, not only the maximum for computing the fabric tensor. They are computed as:

$$\mathsf{SVD} = \int_{\mathbf{v}\in\Omega} \left(\sum_{i=1}^{N} L_i^3(\mathbf{v}) \ \mathbf{v}\,\mathbf{v}^T \right) d\Omega , \tag{8}$$

$$\mathsf{SLD} = \int_{\mathbf{v}\in\Omega} \left(\sum_{i=1}^{N} L_i(\mathbf{v}) \ \mathbf{v}\,\mathbf{v}^T \right) d\Omega , \tag{9}$$

where Ω is the unitary sphere, and $L_i(\mathbf{v})$ is the intercept at i with respect to the orientation $\mathbf{v}$. Thus, the main difference between SVD and SLD is the power of L used in the formulation.

A related fabric tensor is the tensor scale [82, 99]. In this case, every intercepting segment is symmetrized with respect to the reference point by replicating the closest length to the edge in the opposite direction. A local tensor at the sampling positions is computed with the length of the symmetrized lines and the final fabric tensor is computed by adding all local tensors.

In [71] is reported that SVD and SLD are better predictors of mechanical orientation. However, the same study also reports that the MIL is a better predictor

of mechanical anisotropy. Regarding the tensor scale, an initial study reported good correlations between this tensor and mechanical properties [52].

The most important drawback of the methods presented in this subsection is their computational cost. Since these methods are based on a sampling procedure, the accuracy of the computations is related to the complexity of the algorithms. Usually, a huge amount of tests is required to obtain a reasonable accuracy.

3.2.2 Inertia-Based Methods

An straightforward way to compute a volume-based fabric tensor is to compute a global inertia tensor [35] of the material of interest, which is given by:

$$\mathsf{IT} = \int_{\mathbf{x}\in I} \rho(\mathbf{x}) \left((\mathbf{x}-\mathbf{s})^2 \, \mathsf{I} - (\mathbf{x}-\mathbf{s})(\mathbf{x}-\mathbf{s})^T \right) dI \tag{10}$$

where I is the image, $\rho(\mathbf{x})$ is the gray-scale value at $\mathbf{x}$, which is assumed to be proportional to the mass at that point, and $\mathbf{s}$ is the center of mass. However, poor correlations with the MIL tensor have been reported [96], and consequently, it is expected to be a bad predictor of mechanical properties. A possible hypothesis for this is that the path that joins every position to the center of mass usually includes large regions of bone marrow and this fact can influence its appropriateness as fabric tensor.

A possible way to tackle this problem is to compute local inertia tensors computed in local spherical neighborhoods, as proposed in [90], and then to generate a global inertia tensor by adding them up or using any other homogenization scheme. A related strategy is the sampling sphere orientation distribution [89], which adds the gray-scale values of spherical neighborhoods located at some specific sampling locations into a spherical container, as shown in Fig. 5. These neighborhoods are chosen in such a way that their centers are as close as possible to the skeleton of the material of interest. The resulting container is approximated through tensors following an adapted version of the technique proposed in [38]. Since these approximations are related to the computation of the inertia tensor in the container, the method can be seen as a homogenization scheme for computing a global inertia tensor from local inertia information. From the results presented in [89], the use of local inertia tensors partially solves the problems of the global inertia tensor, since the resulting tensors are more correlated with the MIL tensor.

3.3 Texture-Based Methods

The following subsections describe some methods that make use of texture analysis tools to compute fabric tensors.

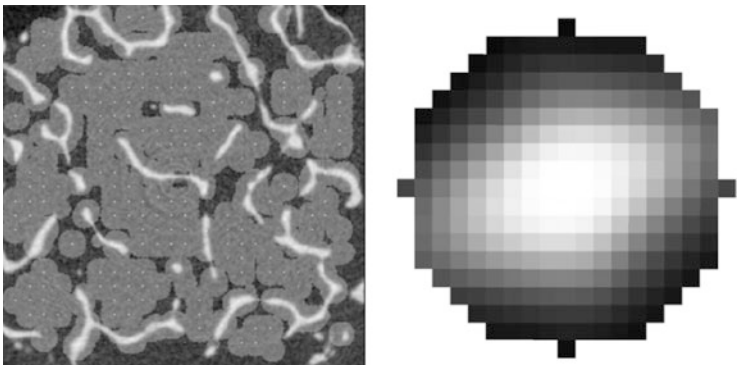


Fig. 5 SSOD. *Left*: the image is sampled with some spheres. *Right*: the gray-scale values are accumulated in a spherical container. Fabric tensors approximate the gray-scale values in the container (Reprinted from [89] with permission from Elsevier)

3.3.1 Fractal-Based Methods

These methods assume a fractal nature of trabecular bone. The basic idea of this approach is to perform directional measurements of fractal dimension (FD) to create orientation distributions that, afterwards, are approximated through tensors. The FD can be computed in many different ways [54]. A basic strategy is the so-called box-counting algorithm where FD is estimated as:

$$\mathrm{FD} = -\lim_{r \to 0} \frac{\log N(r)}{\log(r)} \tag{11}$$

where r is the size of a box and $N(r)$ is the number of boxes required to utterly cover the material of interest. Similar to this method are the skyscrapers and blanket fractal analyses [26].

Alternatively, by assuming a fractional Brownian motion model [5], FD can be computed in the Fourier domain for a specific direction as a function of the slope of the linear regression computed on a log-log plot of the power spectrum vs. frequency [55, 59, 75, 102]. The process is shown in Fig. 6 for a specific orientation.

Since very often this log-log curve does not have a linear behavior for the whole spectrum, it is common to use two fitting lines: for low and high frequencies respectively [55] (see Fig. 6). Some other methods to compute the FD are the augmented Hurst orientation transform [77] and the variance orientation transform [97, 98].

Most of these methods perform the computations in the Fourier domain. Since measurements are performed at specific directions, it is more convenient to sample the Fourier domain by using polar or spherical coordinates instead of Cartesian. However, computing fast Fourier transform in polar/spherical coordinates is not yet a mature technique, although important advances have been done in the last few years [3, 39, 93].

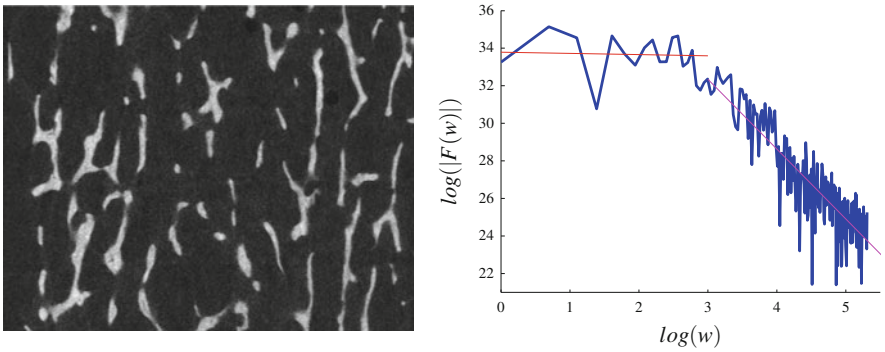


Fig. 6 Estimation of the FD. *Left*: a 2D slice of the image of Fig. 1a. *Right*: log-log plot of the power spectrum vs. frequency at a specific orientation and two linear regressions covering low and high frequencies respectively

A drawback of fractal-based methods is that it is still not clear whether or not trabecular bone follows a fractal pattern with authors in favor [74] and against this hypothesis [10]. From our own experience, the required computation of linear regressions usually involves large errors for images of trabecular bone. As a consequence, since these errors have a direct impact in the computation of fabric tensors, that makes it difficult to obtain reliable and accurate results. Despite this, good correlations with mechanical properties have been reported [55]. Another drawback is that, although the methods can be extended to 3D, they have usually been tested in 2D images of trabecular bone.

3.3.2 Texture Features

Some authors have proposed directional texture features to compute fabric tensors. For example, the line fraction deviation method [20] constructs an orientation distribution from the variance of the gray-scale values along test lines at different orientations, which is then approximated through tensors. Good correlations with the stiffness tensor have been reported for this method [21]. The basic ideas of this method are related to the variance orientation transform from fractal analysis [97, 98].

Related to this strategy, in [92],spatial autocorrelation of the gray-scale values instead of the variance is used to construct the orientation distribution. Their main assumption is that trabecular bone has a quasi-regular structure. However, in [86] is reported that the resulting fabric tensor does not correlate with mechanical properties of trabecular bone. Alternatively, other statistical measurements can be used instead of the variance or spatial autocorrelation to construct the orientation distribution [27].

3.4 Alternative Methods

Recently, the Minkowski tensors have been proposed as an elegant a way to integrate boundary- and volume-based techniques [84, 85]. Six linearly independent Minkowski tensors are defined in 3D:

$$\mathsf{W}_0^{2,0} = \int_{\mathbf{p}\in V} \mathbf{p}\,\mathbf{p}^T\, dV, \tag{12}$$

$$\mathsf{W}_1^{2,0} = \frac{1}{3}\int_{\mathbf{p}\in S} \mathbf{p}\,\mathbf{p}^T\, dS, \tag{13}$$

$$\mathsf{W}_2^{2,0} = \frac{1}{3}\int_{\mathbf{p}\in S} H(\mathbf{p})\,\mathbf{p}\,\mathbf{p}^T\, dS, \tag{14}$$

$$\mathsf{W}_3^{2,0} = \frac{1}{3}\int_{\mathbf{p}\in S} G(\mathbf{p})\,\mathbf{p}\,\mathbf{p}^T\, dS, \tag{15}$$

$$\mathsf{W}_1^{0,2} = \frac{1}{3}\int_{\mathbf{p}\in S} \mathbf{n}\,\mathbf{n}^T\, dS, \tag{16}$$

$$\mathsf{W}_2^{0,2} = \frac{1}{3}\int_{\mathbf{p}\in S} H(\mathbf{p})\,\mathbf{n}\,\mathbf{n}^T\, dS, \tag{17}$$

where $\mathbf{p}$ represents the position of points inside the trabecular bone, V, $\mathbf{n}$ is the normal at $\mathbf{p}$ at the interface between phases, S, and $H(\mathbf{p})$ and $G(\mathbf{p})$ are the mean and Gaussian curvatures at $\mathbf{p}$ respectively. These tensors are called the moment tensor solid, moment tensor hollow, moment tensor wireframe, moment tensor vertices, normal distribution and curvature distribution tensors respectively [85]. Notice that the moment tensor solid and the normal distribution tensor are closely related to the inertia tensor and the GST respectively. Afterwards, different measurement of orientation and anisotropy can be obtained from these six tensors.

Since marrow contains large amounts of water, a promising alternative method to estimate fabric tensors experimentally is through diffusion tensor imaging (DTI) [8, 56, 81]. Although DTI has extensively been used in fiber tractography (see other chapters of this book), its use in trabecular bone is relatively scarce. The following reasons have impeded a faster development of this approach. First, this method computes orientation and anisotropy of bone marrow instead of trabecular bone, so the resulting tensor must conveniently be post-processed in order to obtain a fabric tensor of trabecular bone, and, to our knowledge, such a post-processing has not been proposed so far. Second, it necessary to develop new DTI pulse sequences, since the ones used for white matter in the brain are not appropriate for bone marrow as these two types of tissue have very different magnetic properties and morphology.

A technique used in research of foamy structures is the so-called texture tensor [23]. Given a lattice or mesh, a local texture tensor is computed by aggregating tensorized vectors (i.e. the outer product of these vectors with themselves) that

connect the center of a cell with the centers of neighboring cells. A global texture tensor is then computed by aggregating local texture tensors. A drawback of this technique when it is applied to trabecular bone is that the resulting tensor will depend on the technique used to generate the required input mesh.

An alternative way to construct orientation distributions from the skeleton of trabecular bone was proposed in [41], where mass and thickness of every branch in the skeleton is associated to its orientation. The main drawback of this approach is that it assumes that trabecular bone is composed by rod-like trabeculae, which largely limits its scope of use as it has been shown that this assumption is not always complied [63].

Finally, in [7] anisotropy is directly extracted from a visual examination of the power spectrum of X-ray images. Unlike all methods reviewed in this chapter, this technique is biased by the human observer's perception and, in practice, it can only be used for very anisotropic structures.

4 Relations Between Morphology-Based Fabric Tensors and Mechanics

Morphology-based methods are appealing since they do not have any dependency on boundary conditions, and consequently they are more predictable. However, unlike mechanics-based, morphology-based methods require an extra step of validation with respect to mechanical properties of the tissue, since the quality of a fabric tensor is given by its capacity of predicting mechanical properties in realistic scenarios. Usually, this assessment is performed with respect to a model [103]. A complete review of the models proposed in the literature is presented in [103]. For illustration, two of such models are presented below.

Let $\otimes$ and $\overline{\otimes}$ be the tensorial and double tensorial products of second-order tensors respectively, which, using the Einstein summation notation, are given by [104]:

$$\mathsf{A}\otimes\mathsf{B} = \mathsf{A}_{ik}\mathsf{B}_{jl}, \tag{18}$$

$$\mathsf{A}\overline{\otimes}\mathsf{B} = \frac{1}{2}(\mathsf{A}_{ik}\mathsf{B}_{jl} + \mathsf{A}_{il}\mathsf{B}_{jk}). \tag{19}$$

On the one hand, Cowin [13] proposed that the stiffness c and a fabric tensor M should be related through the formula:

$$\mathsf{c}(\nu,\mathsf{M}) = \sum_{a,b=1}^{3}(\lambda_{ab}(\nu,m_a,m_b)\mathsf{M}_a\otimes\mathsf{M}_b) + \sum_{a,b=1,a\neq b}^{3}2(\mu_{ab}(\nu,\mathbf{m}_a,\mathbf{m}_b)\mathsf{M}_a\overline{\otimes}\mathsf{M}_b), \tag{20}$$

where ν is the volume fraction, m_a and $\mathbf{m}_a$ are the eigenvalues and eigenvectors of the fabric tensor M respectively, $\mathsf{M}_a = \mathbf{m}_a \mathbf{m}_a^T$, λ_{ab} and μ_{ab} are unknown functions of ν, m_a and m_b,

Alternatively, the following stiffness-fabric relation was proposed by Zysset and Curnier [104]:

$$\mathsf{c}(\nu, \mathsf{M}) = \sum_{a,b=1}^{3} ((\lambda_0 + 2\mu_0)\nu^{\alpha} m_a^{\beta} m_b^{\beta} \mathsf{M}_a \otimes \mathsf{M}_a) + \sum_{a,b=1, a \neq b}^{3} (2\mu_0 \nu^{\alpha} m_a^{\beta} m_b^{\beta} \mathsf{M}_a \overline{\underline{\otimes}} \mathsf{M}_b) \tag{21}$$

where α and β are constants and λ_0 and μ_0 are the unknown functions.

Once a model is chosen, multilinear regressions are performed in order to estimate the unknown functions (λ_{ab} and μ_{ab} for Cowin's model, or λ_0 and μ_0 for Zysset and Curnier's model) that minimizes the error between the actual stiffness tensor and the one estimated with the model. The reference stiffness tensor can be estimated either through mechanical simulations as described in Sect. 2.1, or through mechanical experiments [103]. Thus, a morphology-based fabric tensor with a small error between the actual and estimated stiffness tensor is preferred, since this indicates that it is more related to mechanical properties of the tissue.

The MIL tensor usually has the better performance in these assessments. An interesting alternative to minimize the error between the reference and estimated stiffness tensor is to include parameters in the computation of the fabric tensor, as we proposed in [64]. This can give more flexibility to the fitting process, resulting in better estimations of the reference stiffness tensor.

Assessing relations between fabric tensors and mechanical properties of trabecular bone is far from easy. First, these comparisons are mainly restricted to in vitro where the reference stiffness tensor can be obtained. On the one hand, it is difficult to compute reliable stiffness tensors from mechanical simulations from low-resolution images. However, the high-resolution images needed for mechanical simulations are not attainable in vivo for practical and radiation protection issues. On the other hand, invasive mechanical measurements in vivo are not reliable, since they are based on many assumptions. Moreover, such measurements and not always possible for every skeletal site [69, 101].

Reference stiffness tensors obtained through mechanical experiments are preferred to those obtained from mechanical simulations as they are more closely related to reality and do not have the aforementioned problems of mechanical simulations. Unfortunately, it is also difficult to design reliable mechanical experiments in vitro. On the one hand, these experiments are exposed to several sources of error [70, 73]. On the other hand, measurements are usually available in only a few directions (very often in a single one), so many experiments have to be conducted for different directions in order to be combined afterwards, a procedure that is prone to errors. Moreover, it is usually unknown the relationship between the main orientation of trabecular bone and the tested orientations.

For these difficulties, many authors validate their methods by making comparisons with the MIL tensor instead of with the stiffness tensor. However, a direct relationship between a new method and the stiffness tensor is necessary when a better performance than the MIL tensor is being reported.

5 Concluding Remarks

This chapter has presented a comprehensive review of techniques for computing fabric tensors. In general, current methods tend to be less manual and more accurate by addressing most of the disadvantages from previous approaches. Despite this, research in fabric tensors is far from mature and many issues need to be tackled.

First, image acquisition of trabecular bone is challenging in vivo due to the size of the trabecular structure. For example, trabecular thickness ranges between 100 and 300 μm depending on the skeletal site [62], while standard magentic resonance imaging (MRI) and computed tomography (CT) scanners offer resolutions of about 500 and 100 μm respectively. That means that a complete trabecula is covered by at most three voxels, making these images prone to partial volume effects. In addition, blurring, artifacts and noise are not uncommon in these type of images. Hence, it is difficult to perform accurate morphological analyses in vivo. For this reason, methods for computing fabric tensors in gray-scale are appealing, since they are not affected by the accuracy of the segmentation process, which is particularly difficult for images acquired in vivo. Also, the quality of the images are expected to be improved in the next few years, especially with high-resolution peripheral quantitative computed tomography (HR-pQCT) and cone beam computed tomography (CBCT) scanners, which are able to obtain spatial resolutions in the order of 80 μm in vivo [4,43] with very low radiation doses, which can range between 3 and 10 μSv for HR-pQCT [17] and between 11 and 77 μSv for CBCT [53] compared to the 3 mSv usually required by high-resolution multi-detector CT (HR-MDCT) scanners. It is important to remark that, in clinical practice, physicians use dual-energy X-ray absorptiometry (DXA) for measuring the bone mineral density. However, this technique is unable to measure differences in the trabecular structure, which has been shown more related to the development of trabecular bone diseases [42]. Other in vivo techniques such as quantitative ultrasound (QUS) [28, 79] and resonance frequency analyzers (RFA) [1, 58] face the same problematic as DXA.

Second, it is important to remark that fabric tensors are not global measurements. Thus, it is possible to obtain fields of fabric tensors where tensors are computed locally with respect to a neighborhood. However, large neighborhoods are usually used, since regions of interest are usually assumed homogeneous. The net result of this is that the resulting tensor field varies slowly in the space. In consequence, a single tensor is usually computed as a representative measurements for a complete region of interest. However, homogeneity for trabecular bone analysis has been questioned by some authors, e.g., [72]. This imposes the problem of determining the appropriate size of neighborhoods. One strategy to tackle this issue is to propose

measurements of homogeneity of neighborhoods. Regarding fabric tensors, an alternative is to assess changes in orientation and anisotropy with respect to the size of the neighborhood used in the computations. To our knowledge, these types of analysis have not been proposed so far.

Third, some authors argue that it is not necessary to perform measurements in 3D, since features extracted from 3D and 2D projections have been shown correlated [36, 95]. However, this point requires more extensive validation with different features and skeletal sites.

Finally, some authors have found that higher-order tensors are necessary at some skeletal sites, e.g. the calcaneus [21, 22]. This should not be a big problem for most methods, since they usually compute orientation distributions that are approximated through tensors, which can be in theory of any kind. Moreover, methods for performing such approximations are well-established [38, 50, 51]. Despite this, an extensive validation of this point is necessary. On the other hand, it is necessary to bear in mind that, assuming linearity, fourth-order would be the highest necessary order for any fabric tensor, since that is the order of the stiffness tensor. Alternatively, other approximation methods can be used instead of tensors in order to model orientation distributions, e.g., spherical harmonics [41].

To summarize, despite large amount of work in the field, and the advances attained in last decades, there are still many unsolved issues in order to use fabric tensors in clinical practice of trabecular bone diseases. We think proposals to tackle these issues will steer the research in the field in the oncoming years.

References

1. Abtahi, J., Tengvall, P., Aspenberg, P.: A bisphosphonate-coating improves the fixation of metal implants in human bone. a randomized trial of dental implants. Bone **50**(5), 1148–1151 (2012)
2. Aula, A., Töyräs, J., Hakulinen, M., Jurvelin, J.: Effect of bone marrow on acoustic properties of trabecular bone – 3d finite difference modeling study. Ultrasound Med. Biol. **35**(2), 308–318 (2009)
3. Averbuch, A., Coifman, R., Donoho, D., Elad, M., Israeli, M.: Fast and accurate polar Fourier transform. Appl. Comput. Harmon. Anal. **21**(2), 145–167 (2006)
4. Bauer, J.S., Link, T.M.: Advances in osteoporosis imaging. Eur. J. Radiol. **71**(3), 440–449 (2009)
5. Benhamou, C.L., Lespessailles, E., Jacquet, G., Harba, R., Jennane, R., Loussot, T., Tourliere, D., Ohley, W.: Fractal organization of trabecular bone images on calcaneus radiographs. J. Bone. Miner. Res. **9**(12), 1909–1918 (1994)
6. Bigün, J., Granlund, G.H.: Optimal orientation detection of linear symmetry. In: Proceedings of the International Conference on Computer Vision (ICCV), London, pp. 433–438 (1987)
7. Brunet-Imbault, B., Lemineur, G., Chappard, C., Harba, R., Benhamou, C.L.: A new anisotropy index on trabecular bone radiographic images using the fast Fourier transform. BMC Med. Imaging **5**, 4 (2005)
8. Capuani, S., Rossi, C., Alesiani, M., Maraviglia, B.: Diffusion tensor imaging to study anisotropy in a particular porous system: the trabecular bone network. Solid State Nucl. Magn. Reson. **28**(2–4), 266–272 (2005)

9. Charalambakis, N.: Homogenization techniques and micromechanics. A survey and perspectives. Appl. Mech. Rev. **63**, 030,803–1–030,803–10 (2010)
10. Chung, H.W., Chu, C.C., Underweiser, M., Wehrli, F.W.: On the fractal nature of trabecular structure. Med. Phys. **21**(10), 1535–1540 (1994)
11. Ciarelli, T.E., Fyhrie, D.P., Schaffler, M.B., Goldstein, S.A.: Variations in three-dimensional cancellous bone architecture of the proximal femur in female hip fractures and in controls. J. Bone Miner. Res. **15**(1), 32–40 (2000)
12. Coussy, O.: Poromechanics, 2nd edn. Wiley, Chichester (2004)
13. Cowin, S.: The relationship between the elasticity tensor and the fabric tensor. Mech. Mater. **4**(2), 137–147 (1985)
14. Cowin, S.C., Cardoso, L.: Fabric dependence of bone ultrasound. Acta Bioeng. Biomech. **12**(2), 3–23 (2010)
15. Cowin, S.C., Cardoso, L.: Fabric dependence of wave propagation in anisotropic porous media. Biomech. Model Mechanobiol. **10**(1), 39–65 (2011)
16. Cowin, S.C., Doty, S.B.: Tissue Mechanics. Springer, New York (2007)
17. Damilakis, J., Adams, J.E., Guglielmi, G., Link, T.M.: Radiation exposure in X-ray-based imaging techniques used in osteoporosis. Eur. Radiol. **20**(11), 2707–2714 (2010)
18. Felsberg, M., Jonsson, E.: Energy tensors: Quadratic, phase invariant image operators. In: Proceedings of the Symposium German Association for Pattern Recognition (DAGM), Vienna. Lecture Notes in Computer Science, vol. 3663, pp. 493–500 (2005)
19. Förstner, W.: A feature based correspondence algorithm for image matching. Int. Arch. Photogramm. Remote Sens. **26**, 150–166 (1986)
20. Geraets, W.G.: Comparison of two methods for measuring orientation. Bone **23**(4), 383–388 (1998)
21. Geraets, W.G.M., van Ruijven, L.J., Verheij, J.G.C., van der Stelt, P.F., van Eijden, T.M.G.J.: Spatial orientation in bone samples and young's modulus. J. Biomech. **41**(10), 2206–2210 (2008)
22. Gomberg, B.R., Saha, P.K., Wehrli, F.W.: Topology-based orientation analysis of trabecular bone networks. Med. Phys. **30**(2), 158–168 (2003)
23. Graner, F., Dollet, B., Raufaste, C., Marmottant, P.: Discrete rearranging disordered patterns, part I: robust statistical tools in two or three dimensions. Eur. Phys. J. E: Soft Matter Biol. Phys. **25**(4), 349–369 (2008)
24. Granlund, G.H., Knutsson, H.: Signal Processing for Computer Vision. Kluwer, Dordrecht (1995)
25. Groemer, H.: Geometric Applications of Fourier Series and Spherical Harmonics. Cambridge University Press, New York (1996)
26. Guggenbuhl, P., Bodic, F., Hamel, L., Baslé, M.F., Chappard, D.: Texture analysis of X-ray radiographs of iliac bone is correlated with bone micro-CT. Osteoporos. Int. **17**(3), 447–454 (2006)
27. Guggenbuhl, P., Chappard, D., Garreau, M., Bansard, J.Y., Chales, G., Rolland, Y.: Reproducibility of CT-based bone texture parameters of cancellous calf bone samples: influence of slice thickness. Eur. J. Radiol. **67**(3), 514–520 (2008)
28. Hakulinen, M.A., Töyräs, J., Saarakkala, S., Hirvonen, J., Kröger, H., Jurvelin, J.S.: Ability of ultrasound backscattering to predict mechanical properties of bovine trabecular bone. Ultrasound Med. Biol. **30**(7), 919–927 (2004)
29. Helnwein, P.: Some remarks on the compressed matrix representation of symmetric second-order and fourth-order tensors. Comput. Methods Appl. Mech. Eng. **190**(22–23), 2753–2770 (2001)
30. Hollister, S., Kikuchi, N.: A comparison of homogenization and standard mechanics analyses for periodic porous composites. Comput. Mech. **10**(2), 73–95 (1992)
31. Hollister, S.J., Brennan, J.M., Kikuchi, N.: A homogenization sampling procedure for calculating trabecular bone effective stiffness and tissue level stress. J. Biomech. **27**(4), 433–444 (1994)

32. Homminga, J., Van Rietbergen, B., Lochmüller, E.M., Weinans, H., Eckstein, F., Huiskes, R.: The osteoporotic vertebral structure is well adapted to the loads of daily life, but not to infrequent "error" loads. Bone **34**(3), 510–516 (2004)
33. ÒHorn, B.K.P.: Extended Gaussian images. Proc. IEEE **72**(12), 1671–1686 (1984)
34. Ilic, S., Hackl, K., Gilbert, R.: Application of the multiscale FEM to the modeling of cancellous bone. Biomech. Model Mechanobiol. **9**(1), 87–102 (2010)
35. Jähne, B.: Digital Image Processing, 6th edn. Springer, Berlin (2005)
36. Jennane, R., Harba, R., Lemineur, G., Bretteil, S., Estrade, A., Benhamou, C.L.: Estimation of the 3D self-similarity parameter of trabecular bone from its 2D projection. Med. Image Anal. **11**(1), 91–98 (2007)
37. Jupp, P.E., Mardia, K.V.: A unified view of the theory of directional statistics, 1975–1988. Int. Stat. Rev. **57**(3), 261–294 (1989)
38. Kanatani, K.I.: Distribution of directional data and fabric tensors. Int. J. Eng. Sci. **22**(2), 149–164 (1984)
39. Keiner, J., Kunis, S., Potts, D.: Using NFFT 3—a software library for various nonequispaced fast Fourier transforms. ACM Trans. Math. Softw. **36**(4), 19:1–19:30 (2009)
40. Kim, C.H., Zhang, H., Mikhail, G., von Stechow, D., Müller, R., Kim, H.S., Guo, X.E.: Effects of thresholding techniques on microCT-based finite element models of trabecular bone. J. Biomech. Eng. **129**(4), 481–486 (2007)
41. Kinney, J.H., Stölken, J.S., Smith, T., Ryaby, J.T., Lane, N.: An orientation distribution function for trabecular bone. Bone **36**(2), 193–201 (2005)
42. Kleerekoper, M., Villanueva, A.R. Stanciu, J.: The role of three-dimensional trabecular microstructure in the pathogenesis of vertebral compression fractures. Calcif. Tissue Int. **37**(6), 594–597 (1985)
43. Klintström, E., Smedby, Ö., Moreno, R., Brismar, T.: Trabecular bone structure parameters from 3D image processing of clinical multi-slice and cone-beam computed tomography data. Skelet. Radiol. **43**(2), 197–204 (2014)
44. Knutsson, H.: Representing local structure using tensors. In: Proceedings of the Scandinavian Conference on Image Analysis (SCIA), Oulu, pp. 244–251 (1989)
45. Knutsson, H., Westin, C.F., Andersson, M.: Representing local structure using tensors II. In: Proceedings of the Scandinavian Conference on Image Analysis (SCIA), Ystad. Lecture Notes in Computer Science, vol. 6688, pp. 545–556 (2011)
46. Köthe, U., Felsberg, M.: Riesz-transforms versus derivatives: On the relationship between the boundary tensor and the energy tensor. In: Scale Space and PDE Methods in Computer Vision, Hofgeismar. Lecture Notes in Computer Science, vol. 3459, pp. 179–191 (2005)
47. Kreider, J.M., Goldstein, S.A.: Trabecular bone mechanical properties in patients with fragility fractures. Clin. Orthop. Relat. Res. **467**(8), 1955–1963 (2009)
48. Launeau, P., Robin, P.Y.F.: Fabric analysis using the intercept method. Tectonophysics **267**(1–4), 91–119 (1996)
49. Launeau, P., Archanjo, C.J., Picard, D., Arbaret, L., Robin, P.Y.F.: Two- and three-dimensional shape fabric analysis by the intercept method in grey levels. Tectonophysics **492**(1–4), 230–239 (2010)
50. Leng, K.D., Yang, Q.: Fabric tensor characterization of tensor-valued directional data: solution, accuracy, and symmetrization. J. Appl. Math. **2012**, 516,060–1–22 (2012)
51. Li, X., Yu, H.: Tensorial characterisation of directional data in micromechanics. Int. J. Solids Struct. **48**(14–15), 2167–2176 (2011)
52. Liu, Y., Saha, P.K., Xu, Z.: Quantitative characterization of trabecular bone micro-architecture using tensor scale and multi-detector CT imaging. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), Nice. Lecture Notes in Computer Science, vol. 7510, pp. 124–131 (2012)
53. Lofthag-Hansen, S.: Cone beam computed tomography radiation dose and image quality assessments. Swed. Dent. J. Suppl. (209), 4–55 (2009)
54. Lopes, R., Betrouni, N.: Fractal and multifractal analysis: a review. Med. Image. Anal. **13**(4), 634–649 (2009)

55. Majumdar, S., Lin, J., Link, T., Millard, J., Augat, P., Ouyang, X., Newitt, D., Gould, R., Kothari, M., Genant, H.: Fractal analysis of radiographs: assessment of trabecular bone structure and prediction of elastic modulus and strength. Med. Phys. **26**(7), 1330–1340 (1999)
56. Manenti, G., Capuani, S., Fanucci, E., Assako, E.P., Masala, S., Sorge, R., Iundusi, R., Tarantino, U., Simonetti, G.: Diffusion tensor imaging and magnetic resonance spectroscopy assessment of cancellous bone quality in femoral neck of healthy, osteopenic and osteoporotic subjects at 3T: preliminary experience. Bone **55**(1), 7–15 (2013)
57. Martin, R.B.: Toward a unifying theory of bone remodeling. Bone **26**(1), 1–6 (2000)
58. Mc Donnell, P., Liebschner, M., Tawackoli, W., Hugh, P.M.: Vibrational testing of trabecular bone architectures using rapid prototype models. Med. Eng. Phys. **31**(1), 108–115 (2009)
59. Millard, J., Augat, P., Link, T.M., Kothari, M., Newitt, D.C., Genant, H.K., Majumdar, S.: Power spectral analysis of vertebral trabecular bone structure from radiographs: orientation dependence and correlation with bone mineral density and mechanical properties. Calcif. Tissue Int. **63**(6), 482–489 (1998)
60. Mizuno, K., Matsukawa, M., Otani, T., Takada, M., Mano, I., Tsujimoto, T.: Effects of structural anisotropy of cancellous bone on speed of ultrasonic fast waves in the bovine femur. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**(7), 1480–1487 (2008)
61. Moakher, M.: On the averaging of symmetric positive-definite tensors. J Elast **82**, 273–296 (2006)
62. Moreno, R., Borga, M., Smedby, O.: Estimation of trabecular thickness in gray-scale images through granulometric analysis. In: Proceedings of the SPIE Medical Imaging Conference 2012: Image Processing (SPIE), San Diego, vol. 8314, pp. 831451-1–831451-9 (2012)
63. Moreno, R., Borga, M., Smedby, Ö.: Evaluation of the plate-rod model assumption of trabecular bone. In: Proc of the International Symposium on Biomedical Imaging (ISBI), Barcelona, pp. 470–473 (2012)
64. Moreno, R., Borga, M., Smedby, Ö.: Generalizing the mean intercept length tensor for gray-level images. Med. Phys. **39**(7), 4599–4612 (2012)
65. Moreno, R., Borga, M., Klintström, E., Brismar, T., Smedby, Ö.: Correlations between fabric tensors computed on cone beam and microcomputed tomography images. In: Computational Vision and Medical Image Processing IV: VIPIMAGE 2013, Funchal. CRC, pp. 393–398 (2013)
66. Moreno, R., Pizarro, L., Burgeth, B., Weickert, J., Garcia, M.A., Puig, D.: Adaptation of tensor voting to image structure estimation. In: Laidlaw, D., Vilanova, A. (eds.) New Developments in the Visualization and Processing of Tensor Fields, pp. 29–50. Springer, Berlin (2012)
67. Mulvihill, B.M., Prendergast, P.J.: Mechanobiological regulation of the remodelling cycle in trabecular bone and possible biomechanical pathways for osteoporosis. Clin. Biomech. **25**(5), 491–498 (2010)
68. Naili, S., van Rietbergen, B., Sansalone, V., Taylor, D.: Bone remodeling. J. Mech. Behav. Biomed. Mater. **4**(6), 827–828 (2011)
69. Nazer, R.A., Lanovaz, J., Kawalilak, C., Johnston, J.D., Kontulainen, S.: Direct in vivo strain measurements in human bone-a systematic literature review. J. Biomech. **45**(1), 27–40 (2012)
70. Odgaard, A.: Three-dimensional methods for quantification of cancellous bone architecture. Bone **20**(4), 315–328 (1997)
71. Odgaard, A., Kabel, J., van Rietbergen, B., Dalstra, M., Huiskes, R.: Fabric and elastic principal directions of cancellous bone are closely related. J. Biomech. **30**(5), 487–495 (1997)
72. Ostoja-Starzewski, M.: Material spatial randomness: from statistical to representative volume element. Probab. Eng. Mech. **21**(2), 112–132 (2006)
73. Pahr, D.H., Zysset, P.K.: Influence of boundary conditions on computed apparent elastic properties of cancellous bone. Biomech. Model Mechanobiol. **7**(6), 463–476 (2008)
74. Parkinson, I.H., Fazzalari, N.L.: Methodological principles for fractal analysis of trabecular bone. J. Microsc. **198**(Pt 2), 134–142 (2000)
75. Pentland, A.P.: Fractal-based description of natural scenes. IEEE Trans. Pattern Anal. Mach. Intell. **6**(6), 661–674 (1984)

76. Perrot, C., Chevillotte, F., Panneton, R., Allard, J.F., Lafarge, D.: On the dynamic viscous permeability tensor symmetry. J. Acoust. Soc. Am. **124**(4), EL210–EL217 (2008)
77. Podsiadlo, P., Dahl, L., Englund, M., Lohmander, L.S., Stachowiak, G.W.: Differences in trabecular bone texture between knees with and without radiographic osteoarthritis detected by fractal methods. Osteoarthr. Cartil. **16**(3), 323–329 (2008)
78. Podshivalov, L., Fischer, A., Bar-Yoseph, P.Z.: 3D hierarchical geometric modeling and multiscale FE analysis as a base for individualized medical diagnosis of bone structure. Bone **48**(4), 693–703 (2011)
79. Riekkinen, O., Hakulinen, M., Lammi, M., Jurvelin, J., Kallioniemi, A., Töyräs, J.: Acoustic properties of trabecular bone—relationships to tissue composition. Ultrasound Med. Biol. **33**(9), 1438–1444 (2007)
80. van Rietbergen, B., Odgaard, A., Kabel, J., Huiskes, R.: Direct mechanics assessment of elastic symmetries and properties of trabecular bone architecture. J. Biomech. **29**(12), 1653–1657 (1996)
81. Rossi, C., Capuani, S., Fasano, F., Alesiani, M., Maraviglia, B.: DTI of trabecular bone marrow. Magn. Reson. Imaging **23**(2), 245–248 (2005)
82. Saha, P.K., Wehrli, F.W.: A robust method for measuring trabecular bone orientation anisotropy at in vivo resolution using tensor scale. Pattern Recognit. **37**(9), 1935–1944 (2004)
83. Saltykov, S.A.: Stereometric Metallography, 2nd edn. Metallurgizdat, Moscow (1958)
84. Schröder-Turk, G., Kapfer, S., Breidenbach, B., Beisbart, C., Mecke, K.: Tensorial Minkowski functionals and anisotropy measures for planar patterns. J. Microsc. **238**(1), 57–74 (2010)
85. Schröder-Turk, G.E., Mickel, W., Kapfer, S.C., Klatt, M.A., Schaller, F.M., Hoffmann, M.J.F., Kleppmann, N., Armstrong, P., Inayat, A., Hug, D., Reichelsdorfer, M., Peukert, W., Schwieger, W., Mecke, K.: Minkowski tensor shape analysis of cellular, granular and porous structures. Adv. Mater. **23**(22–23), 2535–2553 (2011)
86. Tabor, Z.: On the equivalence of two methods of determining fabric tensor. Med. Eng. Phys. **31**(10), 1313–1322 (2009)
87. Tabor, Z.: Equivalence of mean intercept length and gradient fabric tensors – 3d study. Med. Eng. Phys. **34**(5), 598–604 (2012)
88. Tabor, Z., Rokita, E.: Quantifying anisotropy of trabecular bone from gray-level images. Bone **40**(4), 966–972 (2007)
89. Varga, P., Zysset, P.K.: Sampling sphere orientation distribution: an efficient method to quantify trabecular bone fabric on grayscale images. Med. Image. Anal. **13**(3), 530–541 (2009)
90. Vasilić, B., Rajapakse, C.S., Wehrli, F.W.: Classification of trabeculae into three-dimensional rodlike and platelike structures via local inertial anisotropy. Med. Phys. **36**(7), 3280–3291 (2009)
91. Voyiadjis, G.Z., Kattan, P.I.: Advances in Damage Mechanics: Metals and Metal Matrix Composites with an Introduction to Fabric Tensors. Elsevier, Oxford (2006)
92. Wald, M.J., Vasilić, B., Saha, P.K., Wehrli, F.W.: Spatial autocorrelation and mean intercept length analysis of trabecular bone anisotropy applied to in vivo magnetic resonance imaging. Med. Phys. **34**(3), 1110–1120 (2007)
93. Wang, Q., Ronneberger, O., Burkhardt, H.: Rotational invariance based on Fourier analysis in polar and spherical coordinates. IEEE Trans. Pattern Anal. Mach. Intell. **31**(9), 1715–1722 (2009)
94. Whitehouse, W.J.: The quantitative morphology of anisotropic trabecular bone. J. Microsc. **101**(2), 153–168 (1974)
95. Winzenrieth, R., Michelet, F., Hans, D.: Three-dimensional (3D) microarchitecture correlations with 2D projection image gray-level variations assessed by trabecular bone score using high-resolution computed tomographic acquisitions: effects of resolution and noise. J. Clin. Densitom. **16**(3), 287–296 (2013)
96. Wolfram, U., Schmitz, B., Heuer, F., Reinehr, M., Wilke, H.J.: Vertebral trabecular main direction can be determined from clinical CT datasets using the gradient structure tensor and not the inertia tensor–a case study. J. Biomech. **42**(10), 1390–1396 (2009)

97. Wolski, M., Podsiadlo, P., Stachowiak, G.W.: Directional fractal signature analysis of trabecular bone: evaluation of different methods to detect early osteoarthritis in knee radiographs. Proc. Inst. Mech. Eng. H. **223**(2), 211–236 (2009)
98. Wolski, M., Podsiadlo, P., Stachowiak, G., Lohmander, L., Englund, M.: Differences in trabecular bone texture between knees with and without radiographic osteoarthritis detected by directional fractal signature method. Osteoarthr. Cartil. **18**(5), 684–90 (2010)
99. Xu, Z., Saha, P.K., Dasgupta, S.: Tensor scale: An analytic approach with efficient computation and applications. Comput. Vis. Image. Underst. **116**(10), 1060–1075 (2012)
100. Yang, G., Kabel, J., van Rietbergen, B., Odgaard, A., Huiskes, R., Cowin, S.C.: The anisotropic hooke's law for cancellous bone and wood. J. Elast. **53**(2), 125–146 (1998)
101. Yang, P.F., Brüggemann, G.P., Rittweger, J.: What do we currently know from in vivo bone strain measurements in humans? J. Musculoskelet. Neuronal. Interact. **11**(1), 8–20 (2011)
102. Yi, W.J., Heo, M.S., Lee, S.S., Choi, S.C., Huh, K.H.: Comparison of trabecular bone anisotropies based on fractal dimensions and mean intercept length determined by principal axes of inertia. Med. Biol. Eng. Comput. **45**(4), 357–364 (2007)
103. Zysset, P.K.: A review of morphology-elasticity relationships in human trabecular bone: theories and experiments. J. Biomech. **36**(10), 1469–1485 (2003)
104. Zysset, P., Curnier, A.: An alternative model for anisotropic elasticity based on fabric tensors. Mech. Mater. **21**(4), 243–250 (1995)

Part V
Applications of Tensor Processing

Tensors in Geometry Processing

Eugene Zhang

Abstract Tensor fields have a wide range of applications outside scientific visualization. In this chapter, we review various types of tensors used in geometry processing, including their properties, application requirements, as well as theoretical and practical results. We will focus on the metric tensor and the curvature tensor, two of the most studied tensors in geometry processing.

1 Introduction

Tensor fields have been a major research topic in scientific visualization and medical imaging, due to their wide applicability in physics, chemistry, and biology. Examples of tensor fields in these domains include stress and strain tensors in solid mechanics, velocity gradient tensors in fluid dynamics, and diffusion tensors in medical imaging. In contrast, considerably less research effort has been given to tensors in geometry processing.

Fortunately, increasing attention has been given to tensor fields by the geometry processing community in recent years. Results in tensor field analysis and visualization have been borrowed from communities that traditionally deal with tensors (scientific visualization, medical imaging) and applied to geometry processing applications such as non-photorealistic rendering, surface parameterization, and geometry remeshing.

In differential geometry, there are three tensor fields describing the geometry of the surface. They are termed the first, second, and third *fundamental forms*, and are usually denoted by I, II, and III, respectively. However, these three forms are *not* independent as they are related by an Eq. [8]. Consequently, one often focuses on

E. Zhang (✉)
School of Electrical Engineering and Computer Science, 2111 Kelley Engineering Center, Oregon State University, Corvallis, OR 97331, USA
e-mail: zhange@eecs.oregonstate.edu

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_13, © Springer-Verlag Berlin Heidelberg 2014

the first two fundamental forms, which can be expressed in the language of *metric tensor* and *curvature tensor*, respectively. In the next sections, we will describe these tensors and their applications in geometry processing.

2 Metric Tensor

In this section we describe applications in geometry processing related to the metric tensor.

In the early days of computer graphics when processing speed and memory of computers are much lower than their counterparts today, modeling 3D geometry with complex details was often considered computationally prohibitive. Yet, the ability to render such details is essential to the realism of the synthesized images. To deal with this challenge, the idea of *texture maps* was employed. Basically, the geometry of a surface is modeled with two complementary components. The first component is a surface that approximates the target surface but has much lower geometric details. The second component is a texture map which is an image that contains fine geometry details. When *wrapping* the texture onto the surface appropriately, the resulting rendering has perceptually similar visual quality to images generated by directly modeling the surface with high geometry detail. Thanks to the hardware setup, the texture map approach is much faster than the geometry-only approach. The idea of using texture maps in representing high geometry details has inspired much research in image-based rendering [17].

To connect the two components in this approach, i.e., the surface S and the texture map I (represented as an 2D image), a correspondence between the two must be established. This correspondence, i.e., a map τ from S to I is referred to as the *surface parameterization* for S. One example of this is the world map for the Earth.

Generating a high-quality parameterization given an arbitrary surface is both important and challenging. For example, due to topological constraints, the surface must be cut open in order to be flattened [7]. The curves along which the surface is cut open are referred to as *seams*. Seams require special care as they lead to texture discontinuity when wrapping the texture back to the surface. However, an even more challenging problem is distortion. Distances can be distorted, so can angles and areas. In cartography, such problems can lead to false notions such as that Greenland is larger than Australia and that the South Pole (a point) is a line. In computer graphics, such distortions lead to uneven sampling rates over the surface, since regions of the same area in the surface may be given drastically different areas in the texture map. Consequently, geometry details in regions receiving less-than-average portion of the texture map typically go through more aggressive low-pass filters when the surface signal is stored in the texture map than regions receiving more-than-average portions of the map. This results in greater loss in the details of the surface signal for regions receiving less-than-average areas in the texture map. To ensure the visual realism for regions receiving the smallest share in the texture

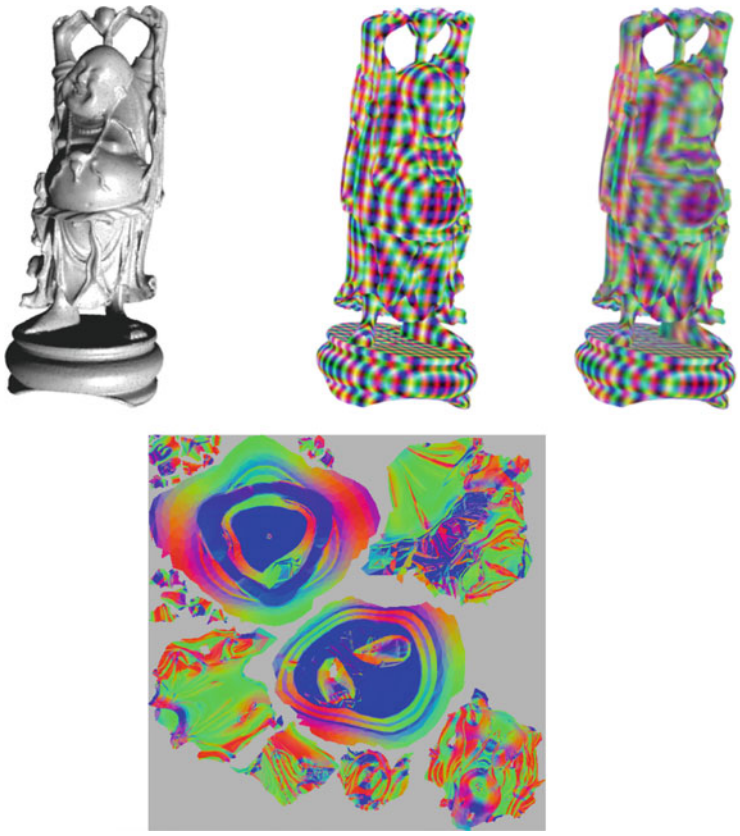


Fig. 1 Distortion in the parameterization (*bottom*) leads to uneven sampling rates on the surface and blurring for regions in the surface that have received less-than-average areas in the texture domain (e.g., Buddha's face and torso). This in turns leads to the loss of details in the texture patterns (*top: right*). Compare it with the true signal (*left*). Notice that the base on which the Buddha stands has a larger-than-average share of the texture map and therefore has preserved finer details in the texture map

domain, one has to use a texture map with a rather large size, which leads to slower performance during rendering. Given that regions in the surface receiving more than a fair share of the parameterization space can already achieve sufficiently high visual quality even with smaller-sized texture maps, increasing the size of the texture map is essentially a waste for such regions. See Fig. 1 for one such example. Next, we will review the cause of distortion and means to reduce distortion.

Consider a surface S represented as a triangular mesh and a parameterization τ that maps every triangle t_i in S to some triangle $\tau(t_i)$ in the plane. Note that sometimes the inverse map of τ, denoted by $\delta = \tau^{-1}$, is referred to as the parameterization for S. We will stay consistent with literature in computer graphics

in which τ is considered the parameterization (also known as texture coordinates). However, our discussion on measuring distortion will be based on δ.

In graphics hardware, a triangle is often the unit of processing. For texture mapping, the signals inside a triangle in the texture domain will be lifted to the corresponding triangle on the surface S through δ as follows:

$$\delta(f_a\tau(v_a) + f_b\tau(v_b) + f_c\tau(v_c)) = f_a(v_a) + f_b(v_b) + f_c(v_c) \tag{1}$$

in which v_a, v_b, and v_c are the vertices of the triangle t in S, and f_a, f_b, and f_c are the barycentric coordinates of points inside t. Notice that δ is piecewise linear which means the distortion is constant over each triangle. To measure distortion, let us assume the total area of the surface is equal to the total area in the texture map. Under this assumption, it is easy to see that a triangle $t \in S$ has a zero stretch only when t is isomorphic to $\tau(t)$. However, when the two triangles are not identical, it is not immediately clear how to measure their difference, i.e., the distortion. For example, the two triangles may be similar (same angle distributions and different areas), or have the same area but different angle distribution, or both. Should we penalize angle distortions more than area distortions, or the other way around? How much of area distortion would be considered equivalent to angle distortion? When mapping t to a similar triangle $\tau(t)$, how to measure distortion when $\tau(t)$ is larger than t and vice versa? All of these questions trace back to a fundamental problem: the distortion is not a scalar, but a tensor.

Let us examine Eq. 1 more closely. Given a triangle $t \in S$, δ_t, the restriction of δ on t, is a bijective linear map. Consider two mutually perpendicular, unit vectors w_1 and w_2 in $\tau(t) \subset \mathbb{R}^2$. The squared length of w_i is given by $|w_i|^2 = w_i \cdot w_i = 1$ for $i = 1, 2$. The angle between the vectors is related to the dot product $w_1 \cdot w_2 = 0$. Finally, the squared area of the parallelogram spanned by w_1 and w_2 is given by $|w_1 \times w_2|^2 = (w_1 \cdot w_1)(w_2 \cdot w_2) - (w_1 \cdot w_2)^2 = 1$.

We are interested in similar quantities for $\delta_t(w_i)$ because the difference in these quantities can give us the distortion in distance, angle, and areas, respectively. Note that δ_t can be represented as a 3×2 matrix M_t that maps a vector w in the plane to a vector $M_t w$ in $\mathbb{R}^3$. The squared length of $\delta(w_i)$ is

$$|\delta_t(w_i)|^2 = (M_t w_i \cdot M_t w_i) = w_i'(M_t' M_t) w_i \tag{2}$$

Notice that the difference between the lengths of w_i and $\delta_t(w_i)$ can only be attributed to the symmetric matrix $G_t = M_t' M_t$. It is straightforward to verify that G_t is also responsible for the angular and areal distortions. G, the tensor field whose restriction to a triangle t is G_t, is referred to as the *metric tensor*. In the ideal setting, i.e., when a triangle $t \in S$ and $\tau(t)$ are isometric, G_t is the identity matrix Id. A parameterization that satisfies this for every triangle is referred to as *isometric parameterization*. However, when distortion is present, how to measure distortion from the metric tensor is a challenging topic. Various measures have been proposed, with respect to which the parameterization algorithms have been optimized. A popular approach is to strive for *conformal parameterizations* which

preserve angles between any pair of vectors. Another possible criterion is *equiareal parameterizations* which preserve areas of the parallelograms spanned by any vector pair. The metric tensor G corresponding to the conformal parameterization and equiareal parameterization satisfies $G_t = k_t Id$ and $det(G_t) = 1$, respectively, for every triangle t. Note that G is isometric if and only if G is both conformal and equiareal. The eigenvalues of G_t correspond to the largest and smallest squared lengths of any unit planar vector under the map δ_t, while the eigenvectors of G_t correspond to directions in which these lengths are achieved. Usually only the eigenvalues are considered important for texture mapping purposes. In terms of the eigenvalues $\sigma_{t,1} \geq 0$ and $\sigma_{t,2} \geq 0$, G_t is isometric, conformal, and equiareal if $\sigma_{t,1} = \sigma_{t,2} = 1$, $\sigma_{t,1} = \sigma_{t,2}$, and $\sigma_{t,1}\sigma_{t,2} = 1$, respectively.

Various measures have been proposed based on $\sigma_{t,1}$ and $\sigma_{t,2}$, essentially providing a tradeoff between conformal and equiareal parameterizations. Example measures include

1. $\sqrt{(\sigma_{t,1} + \sigma_{t,2})(1/\sigma_{t,1} + 1/\sigma_{t,2})}$ [11]
2. $\frac{1}{2}(\sqrt{\sigma_{t,1}} - \sqrt{\sigma_{t,2}})^2$ [6, 20]
3. $\sqrt{\sigma_{t,1} + \sigma_{t,2}}$ [26]
4. $(\sigma_{t,1} - 1)^2 + (\sigma_{t,2} - 1)^2$ [32]

Note that the first two energy formulations from the above consider conformal parameterization as the ideal case, thus ignoring areal distortion. The third energy strikes a balance between conformal and equiareal parameterization. However, in some cases when there are distortions, this energy evaluates to the same as the isometry, such as when $\sigma_{t,1} = 1 + u$ and $\sigma_{t,2} = 1 - u$ for $0 < u < 1$. The fourth energy is minimized if and only if the map is isometric. There are other energy terms not defined directly in terms of the metric tensors, such as the angle-based flattening measure [27] which strive for conformal parameterization.

More recently, there has been work on parameterizing a genus zero triangular mesh surface over a sphere, with applications in remeshing. The general approach is to construct a compatible partition of the mesh surface and the sphere, such as an octahedral partitioning. This allows the sphere and the mesh surface to be mapped onto a common planar domain D. The spherical parameterization is then the composite of the parameterization of the sphere over D and the inverse map of the parameterization of the mesh surface. In this case the metric tensor from the sphere to the mesh surface is also the composition of the metric tensors of the sphere and the mesh surface with respect to D, and the distortions between the sphere and the mesh can again be measured using distortions in the planar parameterizations.

While the original and still primary use of the surface parameterization (and the metric tensor) is in texture mapping, additional applications have been identified, such as fluid simulation on surfaces [28], texture synthesis [24], surface compression [9], and triangular remeshing [2]. With surface parameterization, the computation on mesh surfaces can be transferred into similar (but typically simpler) computations in the plane. The ability to measure distortion in the parameterization is key to achieving desired results as it needs to be reversed when performing the computation in the plane.

Surface parameterization has received much attention from the geometry processing community. However, the focus has been on some derived scalar quantity from the eigenvalues of the metric tensor per triangle. Eigenvectors are typically not considered, and the use of the tensors are isolated, i.e., per triangle. In this sense it is hardly treated as a tensor field, which has structures unique to it. In the next section, we will examine another popular tensor in the geometry processing community, the curvature tensor.

3 Curvature Tensor

The curvature tensor describes the bending of the surface. It has been used in various graphics applications such as non-photorealistic rendering and geometry remeshing.

First, let us consider the curvature of a planar curve γ. Suppose a person is travelling along γ at a constant speed, e.g., the speed limit. Due to the bending of the curve, the person constantly changes his forward direction. However, since he is travelling at a constant speed, the change in his travel direction at any given moment must be perpendicular to the forward direction at the moment. The curvedness of the road can be measured by how sharply the traveler turns. More formally, let $\gamma(s)$ be parameterized by arc length s (equivalent to the milemarkers along the road), the forward travel direction is the tangent to the curve, i.e., $T(s)$. Since the traveler travels at a constant speed $C > 0$, we have $|T(s)|^2 \equiv C^2$. Differentiating both sides with respect to s results in

$$T(s) \cdot N(s) = 0 \tag{3}$$

where $N(s) = T'(s)$ is the normal to the curve. The *signed curvature* at a point $\gamma(s)$ is given by $\kappa(s) = T(s) \times N(s)$.

Let us now consider surfaces. Given a smooth surface $S \subset \mathbb{R}^3$, there are infinitely many curves passing through any point $\mathbf{p} \in S$. Moreover, they may have different curvatures at $\mathbf{p}$. Fortunately results from classical differential geometry state that the curvatures are not a function of individual curves, but of the tangent space at $\mathbf{p}$ [8]. More formally, two curves γ_1 and γ_2 in S have the same curvature at $\mathbf{p}$ if they have the same tangent vectors at $\mathbf{p}$. In addition, given a tangent vector v at $\mathbf{p}$, the curvature corresponding to v is a quadratic function $\kappa(v) = v'Kv$ where K is a symmetric matrix known as the curvature tensor. The eigenvalues of K are referred to as *principal curvatures*, and eigenvectors as *principal directions*. The larger eigenvalue κ_1 is referred to as the *major eigenvalue*, and the smaller eigenvalue κ_2 as the *minor eigenvalue*. Their corresponding eigenvectors are referred to as the major and minor principal directions, respectively.

Unlike the metric tensor which is positive definite, the curvature tensor can have negative eigenvalues. Applying the well-known isotropic-deviatoric decomposition to the diagonalized curvature tensor results in

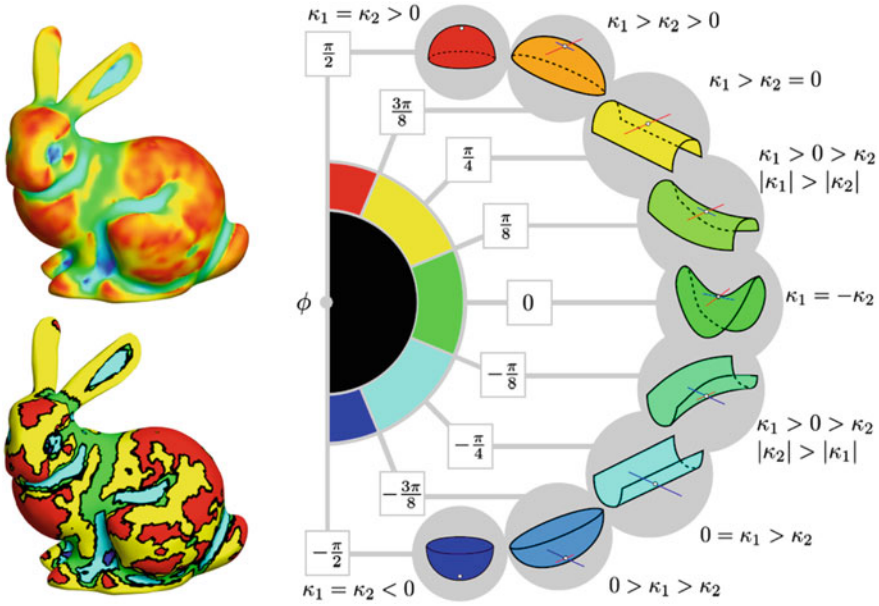


Fig. 2 Surface classification scheme based on the shape index $\phi \in [\pi/2, \pi/2]$ is color mapped to the (*blue, red*) arc in HSV color space: *Left top*: continuous mapping. *Bottom*: binned classification. The legend (*right*) shows surfaces patches which are locally similar to points with given values (This figure is a courtesy of [22], ©2012 IEEE)

$$\begin{pmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{pmatrix} = \frac{\kappa_1 + \kappa_2}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{\kappa_1 - \kappa_2}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{4}$$

Let us consider the vector $\begin{pmatrix} \frac{\kappa_1 - \kappa_2}{2} \\ \frac{\kappa_1 + \kappa_2}{2} \end{pmatrix}$ and write it in the polar form:

$$\rho = \sqrt{(\frac{\kappa_1 - \kappa_2}{2})^2 + (\frac{\kappa_1 + \kappa_2}{2})^2} = \frac{\sqrt{\kappa_1^2 + \kappa_2^2}}{\sqrt{2}} \tag{5}$$

$$\phi = \tan^{-1}(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}) \tag{6}$$

ρ is the tensor magnitude of the curvature tensor and is zero only when the surface is locally planar. It is referred to as the *curvedness*. Recall that $\kappa_1 \geq \kappa_2$. Consequently, ϕ is well defined and satisfies $-\pi/2 \leq \theta \leq \pi/2$. It is referred to as the *shape index* [18]. Figure 2 illustrates the power of this decomposition with the classification over the bunny surface. $H = \frac{\kappa_1 + \kappa_2}{2}$ is referred to as the *mean curvature*, which is not only the average between the principal curvatures, but the average of the curvatures corresponding to the set of all unit tangent vectors.

The quantity $\frac{\kappa_1-\kappa_2}{2}$ shows the anisotropy in the curvature. It achieves minimal value only in spherical regions in the surface. Note that in the discussion of the shape index we do not consider planar regions since it is where the curvedness is zero and can therefore be considered as having any shape index. The quantity $G = (\frac{\kappa_1+\kappa_2}{2})^2-(\frac{\kappa_1-\kappa_2}{2})^2 = \kappa_1\kappa_2$ measures the relative strength between the isotropic and anisotropic parts of the curvature tensor. It is referred as the *Gaussian curvature*. If positive, the point of interest is more isotropic than it is anisotropic, i.e., elliptical. If negative, the point is more anisotropic, i.e., hyperbolic. When zero, the point is cylindrical. Results from classical differential geometry states that the total Gaussian curvature over a closed two-dimensional manifold S with g handles is $2\pi\chi(S)$, where $\chi(S) = 2 - 2g$ is the *Euler characteristic* of S [8].

We now return to the discussion of the eigenvectors of the curvature tensor. Recall that the eigenvectors of the metric tensor do not play a prominent role in surface parameterization. This is not the case for the curvature tensor in graphics, as the principal curvature directions are important geometric characterization of the surface. In pen-and-ink sketching, Interrante [12] has shown that curves following the principal curvature directions can better illustrate a shape in visualization. Moreover, artists typically draw hatches along principal curvature directions despite not having necessary mathematical background in differential geometry.

We define a *major hyperstreamline* as a curve whose tangent coincides with the major principal curvature directions everywhere along its path. A *minor hyperstreamline* can be defined in a similar fashion. Major and minor hyperstreamlines must intersect perpendicularly, since the major and minor principal directions at a point in the surface are mutually perpendicular. However, eigenvectors are not well-defined at points where $\kappa_1 = \kappa_2$, i.e., spherical points. Such points are referred to as the *umbilical points*, which are the equivalent of singularities in vector fields. An umbilical point can be measured in terms of the local tensor field behavior around it. More specifically, consider an isolated umbilical point $\mathbf{p}_0$ which has a neighborhood inside which no other umbilical points exist. Assume this neighborhood N is a topological disk. When travelling along the boundary of the neighborhood ∂N, the normalized eigenvectors along the curve are also travelling on the Gauss circle S^1. Due to tensor field continuity, when one finishes travelling ∂N once, the eigenvectors must have also travelled the Gauss circle a number of times. However, due to sign ambiguity in the eigenvectors, it is possible that the eigenvectors have travelled only half of the circle instead of the full circle. It can be shown that in general the eigenvectors must have travelled $\frac{L}{2}$ times around the Gauss circle where L is an integer. Moreover, L is independent of the size and shape of the neighborhood N as long as N does not contain additional umbilical points beyond $\mathbf{p}_0$ in its interior or on its boundary. Consequently, $\frac{L}{2}$ is considered as the *index* of $\mathbf{p}_0$ and is denoted by $I(\mathbf{p}_0)$. Note that the index is zero if and only if $\mathbf{p}_0$ is not an umbilical point.

The two most fundamental types of umbilical points are wedges (index $1/2$) and trisectors (index $-1/2$). Interestingly, they correspond to the two simplest ways of reversing travel directions when driving a car: U-turn (wedge) and three-point-turn

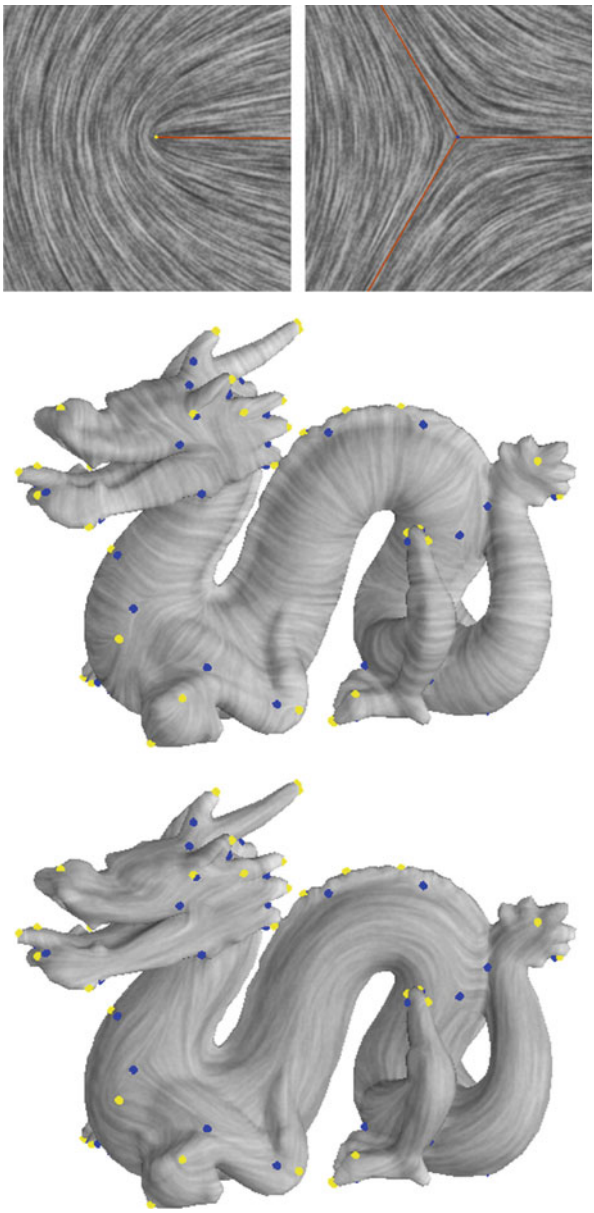


Fig. 3 Two most fundamental umbilical points in the curvature tensor: (*top-left*) wedge, and (*top-right*) trisector. Umbilical points appear in natural locations in shapes (*middle* and *bottom*): wedges in *yellow* and trisectors in *blue*. Shown in the *bottom* are also the major hyperstreamlines (*middle*) and minor hyperstreamlines (*bottom*)

(trisector). See Fig. 3 (top). In addition, umbilical points appear in natural locations in the surface (Fig. 3 (bottom)).

Given a closed, two-dimensional manifold S whose umbilical points are all isolated, Delmarcelle and Hesselink [5] show that

$$2\pi \sum_{J} I(\mathbf{p}_i) = \chi(S) \tag{7}$$

where J is the set of the umbilical points in S. It is interesting to note that the distribution of two seemingly unrelated quantities from the curvature tensor, i.e., the Gaussian curvature and the index of the umbilical points, are both constrained by the topology of the underlying surface. The curvature tensor can be computed using a number of methods [3, 21, 23, 25]. We refer interested readers to these papers for details.

Next we consider some graphics and geometry applications in which the curvature tensor plays a prominent role.

3.1 Non-photorealistic Rendering

Pen-and-ink sketching is a well-researched topic in Non-Photorealistic Rendering (NPR). In a typical setting, a set of lines (hatches), usually monochromatic (e.g., black) are placed against a background (typically white). The locations and densities of the lines are used to outline the shapes, present the main features in the objects, and convey the shading. An NPR system must determine the location, orientation, and density of hatches. Most existing automatic hatching algorithms differ in how they extract some or all of this information.

3.1.1 Feature Line Drawing

The first class of algorithms extract line features from the shapes and highlight these lines. These classes of algorithms have direct application in engineering illustration and CAD and medical data visualization. Consequently, the lines are typically chosen to reflect the underlying geometry without the effect of shading. The difference among this class of algorithms lies in how line features are defined and extracted. We will review some of line definitions that are most relevant to the curvature tensor.

The most commonly used lines in line drawing are perhaps *visible contours*. Line drawing based only on visible contours is sometimes referred to as *silhouette drawing*. Notice that contours are view-dependent. When the viewpoint is changed, the set of contour points is also changed. However, intrinsic features in a mesh, like sharp edges, are not guaranteed to be part of the contour for any appropriate viewpoint. This has led to view-independent line features, such as ridges and valleys which can be defined in terms of the curvature tensor.

A point $\mathbf{p} \in S$ is a *ridge point* if the following conditions are met:

1. $\kappa_1 > |\kappa_2|$, i.e., the absolutely maximal bending occurs in the major principal curvature direction.

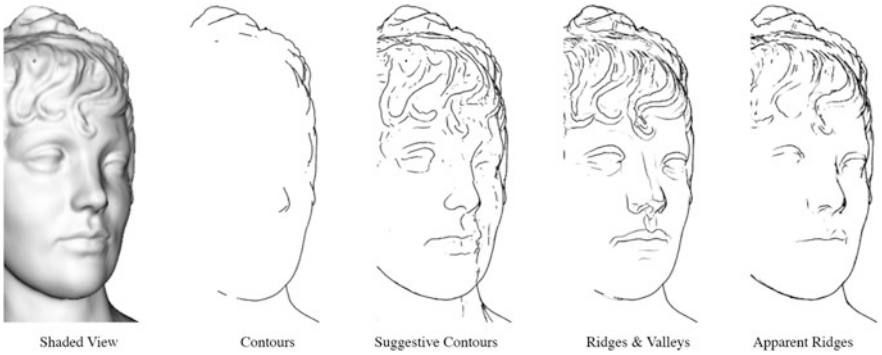


Fig. 4 A comparison of a number of feature-based drawing techniques (This figure is a courtesy of [14], ©2007 ACM)

2. $\mathbf{p}$ is a local maxima of κ_1 on the unique major hyperstreamline containing $\mathbf{p}$

Similarly, a point $\mathbf{p} \in S$ is a *valley point* if

1. $-\kappa_2 > |\kappa_1|$, i.e., the absolutely maximal bending occurs in the minor principal curvature direction, and
2. $\mathbf{p}$ is a local minima of κ_2 on the unique minor hyperstreamline containing $\mathbf{p}$

We will omit the details for computing ridges and valleys here and instead refer interested readers to [23, 25] (Fig. 4).

It has been noted that visible contours often do not produce sufficient lines that reveal the underlying geometry. On the other hand, ridges and valleys are not view-independent. To address these difficulties, the concepts of *suggestive contours* [4], *apparent ridges* [14], and *demarcating curves* [19] are introduced. Both suggestive contours and apparent ridges are view-dependent, while demarcating curves are view-independent.

Roughly speaking, points on suggestive contours given a viewpoint V are not part of the contours with respect to V, but contours with respect to some nearby viewpoint. More formally, a point $\mathbf{p} \in S$ is on the suggestive contour with respect to viewpoint V if

$$D_{\mathbf{w}}\kappa_r > 0 \tag{8}$$

where $\mathbf{w}$ is the projection of $V - \mathbf{p}$ onto the tangent plane at $\mathbf{p}$, κ_r is the curvature at $\mathbf{p}$ in the direction $\mathbf{w}$, and $D_{\mathbf{w}}\kappa_r$ is the derivative of κ_r in the direction of $\mathbf{w}$. Equivalently, a point is on the suggestive contour if it is a local minima of $N \cdot (V - \mathbf{p})$. The computation of suggestive contours can be performed in both object-space and image-space, leading to different tradeoffs between accuracy and numerical stability.

Another view-dependent feature measure is *apparent ridges*. Apparent ridges differ from ridges as the latter is derived from the curvature tensor, while the former

from the *projected curvature tensor* onto the image space. Due to the distortion in orthographic and perspective projections, the projected curvature tensor differs from the curvature tensor. Apparent ridges are therefore the ridges extracted from the projected curvature tensor. One of the main motivations behind apparent ridges is to capture places where luminance would change rapidly should the model be shaded. This notion is in a way related to the idea of suggestive contours as the latter also tries to capture lines that are not features in the current but nearby viewpoints.

Finally, demarcating curves are considered in addition to ridges and valleys. Demarcating curves are transition points between ridges and valleys, much in the same sense as that *inflection points* of a function f ($f'' = 0$) are the transition between local maxima ($f' = 0$ and $f'' < 0$) and local minima ($f' = 0$ and $f'' > 0$). More formally, the gradient of the curvature tensor, C_{ijk}, is a third-order tensor. Given $v = v^i$, a unit tangent vector, $C_{ijk}v^i v^j v^k$ gives the rate of curvature change in v. The direction in which this change is the largest is defined as the *direction for maximal normal curvature variation*. A point $\mathbf{p}$ is on the demarcating curve if $v_{\mathbf{p}}^T K v_{\mathbf{p}} = 0$ where $v_{\mathbf{p}}$ is the direction for maximal normal curvature variation at $\mathbf{p}$.

While feature lines have been a powerful tool in illustrating shapes, they are not often well-suited for surfaces that lack lines features, such as smooth surfaces like a cylinder. In addition, feature-based line drawing primarily aims to revealing geometric features rather than shading effects. Hatching is a more suitable alternative in these cases, which we review next.

3.1.2 Illustration of Smooth Surfaces

Drawing as a form of art often captures our attention in terms of the shading effect and varying geometric details, even when there are few geometric features (such as ridges and valleys) present in the shape. For example, consider a cylinder of an infinite height. Line features as described in Sect. 3.1.1 can only capture the silhouette of the cylinder, i.e., two straight lines. It would be difficult, without drawing the caps of the cylinder, to realize that it is a cylinder.

Hatching provides a nice alternative way of illustrating the shapes in this case. Hatches are used to present the shading effect as well as the internal bending in the geometry, thus making it possible to understand the geometry of an otherwise featureless surface. The key issues for hatching are:

1. How to use hatches to present geometric details?
2. How to use hatches to provide shading effects?

Interrante [12] shows that when hatches follow the principal curvature directions, the shape is best illustrated. Moreover, the lighting effect can be simulated with the density of hatches. Denser hatches indicates lower luminance, while lower or no hatches indicates bright spots or highlight under the viewing condition. Cross-hatching, i.e., drawing two families of mutually perpendicular lines can increase the darkness of a region without increasing the density of the hatches. Consequently,

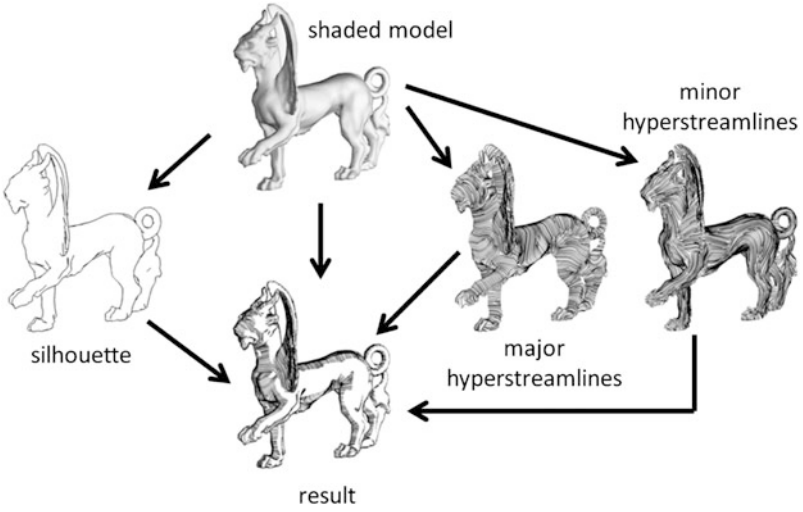


Fig. 5 The process of generating a hatch-based drawing from an input model

cross-hatched regions are often reserved for deep shadows, while single-hatched regions (only one family of hatches) are used for regions that are neither in deep shadows nor in the highlight (Fig. 5).

As mentioned earlier, the principal curvature directions can be computed using a number of methods [3, 21, 23, 25]. Once the curvature tensor field has been computed, a family of evenly spaced hyperstreamlines can be generated for the major principal curvature field and the minor principal curvature field, respectively. Generating evenly spaced hyperstreamlines can be achieved by adapting techniques generating evenly spaced streamlines in vector fields [13]. These two families of hyperstreamlines will be used to generate two images, one for the major and the other for the minor hyperstreamlines. We refer to these two images as I_1 and I_2, respectively. In addition, an image I_3 based on visible contours are also generated. Furthermore, a pixel with a value of 1 is white and a value of 0 is black. These three images will then be composed into a single image as follows:

$$I(p) = \begin{cases} 0 & \text{if } I_3(p)=0 \\ 1 & \text{if in highlight} \\ min(I_1(p), I_2(p)) & \text{if in shadow} \\ I_2(p) & \text{otherwise} \end{cases} \tag{9}$$

Note that in the above one can also choose to always use $I_1(p)$ for single-hatched regions. Another means of generating I_1 and I_2 is to first project the principal directions onto the image plane and trace hyperstreamlines in the image plane. This alternative is view-dependent but typically is fast enough for interactive applications. In contrast, the object-based approach requires much time for

pre-processing but is then ready for interactive display, except when the resolution is changed.

3.2 Quadrangular and Triangular Remeshing

The curvature tensor has also been used in geometry remeshing, which refers to generating a new mesh from an input mesh (typically triangular) subject to some optimization criteria. If the output mesh is also triangular, the process is referred to as *triangular remeshing*. On the other hand, if the output is a quad mesh, the process is referred to as *quadrangular remeshing*.

Classical triangular remeshing approaches place a set of points on the input mesh surface according to some density constraints [2,10,29,30]. These points will be the vertices of the output mesh. The density of the points is usually required to reflect the geometry in the input. Consequently, some forms of curvature (mean curvature, Gaussian curvature, etc.) are considered as part of the density function. Delauney triangulation or centroidal Voronoi tessellation can then be performed on the point set to generate the triangulation.

Alliez et al. [1] revitalize the topic of quadrangular remeshing. In their pioneering work, a quad-dominant mesh is generated by intersecting one family of evenly spaced major hyperstreamlines with one family of evenly spaced minor hyperstreamlines. Since the two families intersect at the right angle, the resulting quad-dominant mesh consists of mostly nice rectangles. Moreover, the edges in the mesh follow the principal curvature directions and therefore have low approximation error.

Due to numerical issues, there are usually an excessive number of umbilical points in the output mesh, leading to a large number of irregular vertices, i.e., the valency is not four. Irregular vertices, especially when misplaced, can lead to difficulties in subsequent mesh processing. Zhang et al. [31] introduce operations to control the number and location of umbilical points by cancelling a pair of umbilical points with opposite tensor indexes, or by moving an umbilical point to a more appropriate location. We refer to [31] for details.

To remove T-junctions, which occur when the tracing of a hyperstreamline stops without reaching an umbilical point, Kälberer et al. [15,16] make use of the mathematical concept of *covering space* and convert the tensor field to a vector field in the covering space. The vector field is then made curl-free through the Hodge decomposition, and the quadrangulation is performed in the covering space which nicely maps to a pure quad mesh (no T-junctions) in the original input mesh.

Nieser et al. [22] use a similar idea (covering space) for triangulation. Unlike quadrangular remeshing, in which both the major and minor principal directions can be used as edges in the remeshed quad mesh, in triangular remeshing at most one can be used. To deal with this they make use of the shape index and ensure that the edges in the triangles will be aligned with the minor eigenvector directions in ridge-like regions and the major eigenvector directions in valley-like regions.

4 Conclusion

Tensor fields are important to not only scientific visualization and medical imaging, but also computer graphics and geometry processing. In this chapter we review two of the most popular tensors, the metric tensor and the curvature tensor, with applications in surface parameterization, non-photorealistic rendering, and remeshing. We expect that more graphics and geometry applications will be identified for tensor fields, and we believe that research in tensor fields can continue to benefit the visualization, image processing, and graphics communities.

References

1. Alliez, P., Cohen-Steiner, D., Devillers, O., Lévy, B., Desbrun, M.: Anisotropic polygonal remeshing. ACM Trans. Graph. (SIGGRAPH 2003) **22**(3), 485–493 (2003)
2. Alliez, P., Meyer, M., Desbrun, M.: Interactive geometry remeshing. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '02, San Antonio, pp. 347–354. ACM, New York (2002). doi:10.1145/566570.566588, http://doi.acm.org/10.1145/566570.566588
3. Cohen-Steiner, D., de Verdière, É.C., Yvinec, M.: Conforming delaunay triangulations in 3d. In: Symposium on Computational Geometry, Barcelona, pp. 199–208 (2002)
4. DeCarlo, D., Finkelstein, A., Rusinkiewicz, S., Santella, A.: Suggestive contours for conveying shape. In: ACM SIGGRAPH 2003 Papers, SIGGRAPH'03, San Deigo, pp. 848–855. ACM, New York (2003). doi:10.1145/1201775.882354, http://doi.acm.org/10.1145/1201775.882354
5. Delmarcelle, T., Hesselink, L.: The topology of symmetric, second-order tensor fields. In: IEEE Visualization Conference, pp. 140–147 (1994)
6. Desbrun, M., Meyer, M., Alliez, P.: Intrinsic parameterizations of surface meshes. Comput. Graph. Forum **21**(3), 209–218 (2002)
7. Floater, M.S., Hormann, K.: Surface parameterization: a tutorial and survey. In: Advances in Multiresolution for Geometric Modelling, pp. 157–186. Springer, Berlin (2005)
8. Gray, A.: Modern Differential Geometry of Curves and Surfaces with Mathematica, 1st edn. CRC, Boca Raton (1996)
9. Gu, X., Gortler, S.J., Hoppe, H.: Geometry images. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH'02, Bristol, pp. 355–361. ACM, New York (2002). doi:10.1145/566570.566589, http://doi.acm.org/10.1145/566570.566589
10. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., Stuetzle, W.: Mesh optimization. In: SIGGRAPH, Anaheim, pp. 19–26 (1993)
11. Hormann, K., Greiner, G.: MIPS: an efficient global parametrization method. In: Laurent, P.J., Sablonnière, P., Schumaker, L.L. (eds.) Curve and Surface Design: Saint-Malo 1999, Innovations in Applied Mathematics, pp. 153–162. Vanderbilt University Press, Nashville (2000)
12. Interrante, V.: Illustrating surface shape in volume data via principal direction-driven 3d line integral convolution. In: SIGGRAPH'97: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, pp. 109–116. ACM/Addison-Wesley, New York (1997)
13. Jobard, B., Lefer, W.: The motion map: efficient computation of steady flow animations. In: IEEE Visualization, Phoenix, pp. 323–328 (1997)

14. Judd, T., Durand, F., Adelson, E.: Apparent ridges for line drawing. In: ACM SIGGRAPH 2007 Papers, SIGGRAPH'07, San Diego. ACM, New York (2007). doi:10.1145/1275808.1276401, http://doi.acm.org/10.1145/1275808.1276401
15. Kälberer, F., Nieser, M., Polthier, K.: Quadcover – surface parameterization using branched coverings. Comput. Graph. Forum **26**(3), 375–384 (2007)
16. Kälberer, F., Nieser, M., Polthier, K.: Stripe parameterization of tubular surfaces. In: Pascucci, V., Hagen, H., Tierny, J., Tricoche, X. (eds.) Topological Methods in Data Analysis and Visualization. Theory, Algorithms, and Applications., Mathematics and Visualization. Springer, Berlin/Heidelberg (2010)
17. Kang, S.B.: A survey of image-based rendering techniques. In: In Videometrics, SPIE, pp. 2–16. Digital, Cambridge Research Laboratory, Cambridge (1999)
18. Koenderink, J.J., van Doorn, A.J.: Surface shape and curvature scales. Image Vision Comput. **10**, 557–565 (1992)
19. Kolomenkin, M., Shimshoni, I., Tal, A.: Demarcating curves for shape illustration. In: ACM SIGGRAPH Asia 2008 papers, SIGGRAPH Asia'08, Singapore, pp. 157:1–157:9. ACM, New York (2008). doi:10.1145/1457515.1409110, http://doi.acm.org/10.1145/1457515.1409110
20. Lévy, B., Petitjean, S., Ray, N., Maillot, J.: Least squares conformal maps for automatic texture atlas generation. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH'02, Bristol, pp. 362–371. ACM, New York (2002). doi:10.1145/566570.566590, http://doi.acm.org/10.1145/566570.566590
21. Meyer, M., Desbrun, M., Schröder, P., Barr, A.H.: Discrete differential-geometry operators for triangulated 2-manifolds. In: Proc. of the Int. Workshop on Visualization and Mathematics, pp. 35–57 (2002)
22. Nieser, M., Palacios, J., Polthier, K., Zhang, E.: Hexagonal global parameterization of arbitrary surfaces. IEEE Trans. Vis. Comput. Graph. **18**(6), 865–878 (2012). doi:10.1109/TVCG.2011.118, http://dx.doi.org/10.1109/TVCG.2011.118
23. Ohtake, Y., Belyaev, A., Seidel, H.P.: Ridge-valley lines on meshes via implicit surface fitting. ACM Trans. Graph. **23**(3), 609–612 (2004). doi:10.1145/1015706.1015768, http://doi.acm.org/10.1145/1015706.1015768
24. Praun, E., Finkelstein, A., Hoppe, H.: Lapped textures. In: SIGGRAPH'00: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, New Orleans, pp. 465–470. ACM/Addison-Wesley, New York (2000)
25. Rusinkiewicz, S.: Estimating curvatures and their derivatives on triangle meshes. In: Proceedings of the 3D Data Processing, Visualization, and Transmission, 2nd International Symposium, 3DPVT'04, Thessaloniki, pp. 486–493. IEEE Computer Society, Washington (2004). doi:10.1109/3DPVT.2004.54, http://dx.doi.org/10.1109/3DPVT.2004.54
26. Sander, P.V., Snyder, J., Gortler, S.J., Hoppe, H.: Texture mapping progressive meshes. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH'01, Los Angeles, pp. 409–416. ACM, New York (2001). doi:10.1145/383259.383307, http://doi.acm.org/10.1145/383259.383307
27. Sheffer, A., Lévy, B., Mogilnitsky, M., Bogomyakov, A.: Abf++: fast and robust angle based flattening. ACM Trans. Graph. **24**(2), 311–330 (2005). doi:10.1145/1061347.1061354, http://doi.acm.org/10.1145/1061347.1061354
28. Stam, J.: Flows on surfaces of arbitrary topology. ACM Trans. Graph. (SIGGRAPH 2003) **22**(3), 724–731 (2003)
29. Turk, G.: Re-tiling polygonal surfaces. In: Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH'92, New York, pp. 55–64. ACM, New York (1992). doi:10.1145/133994.134008, http://doi.acm.org/10.1145/133994.134008
30. Yan, D.M., Lévy, B., Liu, Y., Sun, F., Wang, W.: Isotropic remeshing with fast and exact computation of restricted voronoi diagram. In: Proceedings of the Symposium on Geometry Processing, SGP'09, Berlin, pp. 1445–1454. Eurographics Association, Aire-la-Ville (2009). http://dl.acm.org/citation.cfm?id=1735603.1735629

31. Zhang, E., Hays, J., Turk, G.: Interactive tensor field design and visualization on surfaces. IEEE Trans. Vis. Comput. Graph. **13**(1), 94–107 (2007)
32. Zhang, E., Mischaikow, K., Turk, G.: Feature-based surface parameterization and texture mapping. ACM Trans. Graph. **24**, 1–27 (2005)

Preliminary Findings in Diagnostic Prediction of Schizophrenia Using Diffusion Tensor Imaging

Yogesh Rathi, Martha E. Shenton, and Carl-Fredrik Westin

Abstract We describe a probabilistic technique for diagnostic prediction of first-episode schizophrenia patients based on their brain diffusion MRI data. The method begins by transforming each voxel from a high-dimensional diffusion weighted signal to a low-dimensional diffusion tensor representation. Three orthogonal diffusion measures (fractional anisotropy, norm, mode) that capture different aspects of the local tissue properties are derived from this diffusion tensor representation. Next, we compute a one-dimensional probability density function of each of the diffusion measures with values obtained from the entire brain. This representation is affine invariant, thus obviating the need for registration of the images. We then use a Parzen window classifier to estimate the likelihood of a new patient belonging to either group. To demonstrate the technique, we apply it to the

Y. Rathi (✉)
Laboratory of Mathematics in Imaging, Psychiatry Neuroimaging Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Department of Psychiatry, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
e-mail: yogesh@bwh.harvard.edu

M.E. Shenton
Psychiatry Neuroimaging Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Department of Psychiatry, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
e-mail: shenton@bwh.harvard.edu

C.-F. Westin
Laboratory of Mathematics in Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
e-mail: westin@bwh.harvard.edu

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_14,

analysis of 22 first-episode schizophrenic patients and 20 normal control subjects. With leave-one-out cross validation, we find a detection rate of 90.91 % (10 % false positives). We also provide several error bounds on the performance of the classifier.

1 Introduction

A recent World Health Organization (WHO) report estimates that nearly 11 % of the population world-wide is affected by some form of brain disorder. These illnesses can often be psychologically and financially devastating to patients, their families and the larger community. Nearly 1 % of the population in the US is affected by schizophrenia. A growing body of evidence suggests that the early stages of schizophrenia (and many other brain disorders) are critical in forming and predicting the course and outcome of the disorder. The classification tools proposed in this work can serve as a first step towards early detection of schizophrenia, which may result in a better prognosis and functional outcome.

Both, post-mortem and neuroimaging studies have contributed significantly to what we know about the brain. Moreover, MRI studies of volumetric reduction in several brain regions in schizophrenia have been particularly informative with respect to confirming early speculations that the brain is disordered in schizophrenia. This work is largely the result of advances in neuroimaging that allowed for more careful measurement of regions of interest within the brain (see review in [15, 22]). In particular, a shape based framework was recently developed by the authors in [8], which utilized volumetric differences as discriminatory features for distinguishing a population of schizophrenia from normal controls (NC). Another recent work in this direction was proposed in [20], where the authors used the affine parameters obtained during registration of each subject to a given atlas as a discriminant feature. However, both works, used structural MRI data, as opposed to diffusion MRI being used in this study.

The advent of diffusion magnetic resonance imaging (dMRI) has provided the opportunity for non-invasive investigation of neural architecture of the brain. Using this imaging technique, neuroscientists want to ask how neurons originating from one region connect to other regions, or how well-defined those connections may be. One of the models that is widely used to analyze dMRI images is the diffusion tensor model. Diffusion tensors represent the diffusion of water molecules in three orthogonal directions with the principal direction aligned with the fiber orientation. Several scalar measures derived from this model are used to assess the strength of connectivity in neural fiber bundles.

Experimental evidence suggests that the tissue component responsible for the observed orientational anisotropy in white matter is principally the spatial organization of cellular membranes, which is modulated by the degree of myelinization of the individual axons and the density of cellular packing [6]. As a result, fractional anisotropy (FA) is the most popular measure used to study abnormalities in white matter [15].

While there has been an explosion in the number of studies reporting statistical differences in various regions of the brain (see [15] and the references therein), very few works have addressed the problem of classifying schizophrenic patients from healthy subjects. Apart from structural MRI data being used by Davatzikos et al. [8] and Pohl and Sabuncu [20], there has been work done by Caan et al. [4], where the authors use dimensionality reduction followed by linear discriminant analysis for classification of patients with schizophrenia (chronic). They use the fractional anisotropy and linear anisotropy images derived from single tensor estimation as discriminant features.

In all of the approaches listed above [4,8], the authors first perform a registration of all the subjects to a common atlas space and subsequent analysis is done on this dataset. However, such an analysis is computationally expensive and requires a very good registration algorithm to spatially normalize all the subjects in one common co-ordinate system. In the proposed work, we compute the probability distribution of several diffusion measures over the entire brain white matter, which makes it un-necessary to register the diffusion images, thus reducing computational complexity. Note that, the probability distribution (by its definition) is invariant to affine-transformations (see Sect. 2.2 for more details).

2 Methods

The overall outline of the proposed algorithm for computing the probabilistic representation of each subject is shown in Fig. 1. From the dMRI scan of a subject, diffusion tensors are first estimated. Three orthogonal diffusion measures (fractional anisotropy (FA), norm (N), mode (M_d)) [14] that form the discriminatory features of our classifier are then computed at each voxel in the white matter region. A nonparametric density estimator is then used to convert each of these measures of each subject into a probabilistic representation, which is affine invariant. Note that, we compute a 1D probability distribution function (pdf) of each of the diffusion measures (FA, N, M_d) from values obtained throughout the white matter. These three one-dimensional pdf's form the features for each subject. This representation is subsequently used by a Parzen window classifier to compute the probability of a previously unseen subject being FE or NC in a cross-validation scheme. Details on each of these steps are given in the subsequent sections.

2.1 Preliminaries

In diffusion weighted imaging, image contrast is related to the strength of water diffusion. At each image voxel, diffusion is measured along a set of distinct gradients, $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{S}^2$ (on the unit sphere), producing the corresponding signal, $\mathbf{s} = [s_1, \ldots, s_n]^T \in \mathbb{R}^n$. The diffusion tensor is related to the signal using the following relation [3,16]:

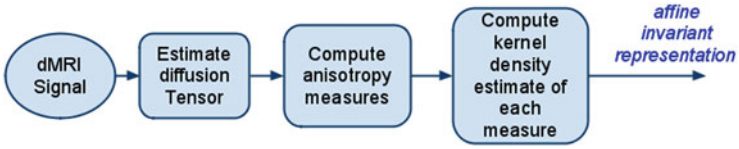


Fig. 1 Overall outline for computing a probabilistic representation for each subject

$$s_i = s_0 \exp(-b\mathbf{u}_i^T D\mathbf{u}_i),$$

where s_0 is a baseline signal intensity, b is an acquisition-specific constant, and D is a tensor describing the diffusion pattern. D can be estimated using a weighted least-squares approach [1].

Several scalar measures derived from the single tensor model have been proposed in the literature [14,18,24]. In particular, we use a set of three orthogonal invariants studied in [14], namely the norm N, fractional anisotropy FA and mode M_d. These measures capture different (orthogonal) aspects of the shape of the tensor. Given, a diffusion tensor D, these measures can be computed as follows:

$$\begin{aligned} N &= \| D \|, \quad FA = \frac{\sqrt{3} \| D - \frac{1}{3} tr(D) I \|}{\sqrt{2} \| D \|}, \\ M_d &= 3\sqrt{6} \left| \left(\frac{\tilde{D}}{\| \tilde{D} \|} \right) \right|, \quad \tilde{D} = D - \frac{1}{3} tr(D) I \end{aligned} \tag{1}$$

where, $|.|$ denotes the determinant, $tr(.)$ is the trace and $\| . \|$ denotes the frobenius norm of a matrix. Thus, FA measures how the shape of the tensor deviates from that of a sphere (diameter of the sphere is given by the average length of the axes of the ellipsoid (tensor)). M_d indicates the mode of the tensor, i.e. $M_d = -1$ indicates planar anisotropy, $M_d = 0$ indicates an orthotropic tensor and $M_d = 1$ indicates linear anisotropic tensor. Norm N measures the "size" of the diffusion tensor. Of these measures, only FA has been widely used to study white matter abnormalities in schizophrenia (see references in [15]). From the above discussion, at voxel $\mathbf{r}$, we compute the following 3-dimensional vector

$$\mathbf{f}(\mathbf{r}) = [N(\mathbf{r}), FA(\mathbf{r}), M_d(\mathbf{r})]^T. \tag{2}$$

2.2 Probabilistic Representations

Probability density functions (PDF) are invariant to translation, rotation, scale and shear of an image, i.e. PDF's are invariant under linear transformation of the coordinates of an image. A nonparametric estimate of the PDF can be computed

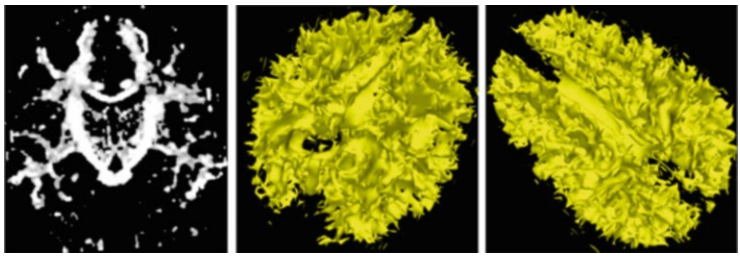


Fig. 2 *Left*: Coronal slice shows region of the brain included in the classifier. This corresponds to FA ≥ 0.4. The other two figures show different views of the volume rendering of the thresholded FA image

using the following expression [19]:

$$p(z) = \frac{1}{Mh} \sum_{i=1}^{M} G\left(\frac{z - I(x)}{h}\right), \quad z \in \{\text{Range of } I\} \tag{3}$$

where $I(x)$ is a scalar image at spatial location x, M is the number of data points, G is a Gaussian kernel and h denotes the bandwidth of the kernel. An affinely transformed image $\tilde{I}$ is related to the original image using the relation $\tilde{I}(Ax) = I(x)$, where A is an affine transformation. Notice that only the co-ordinates of the image I change without changing the image intensities (scalar values). By applying a change of variable in Eq. (3), one can easily see that the PDF $p(z)$ is invariant under affine (linear) transformations.

The proposed set of diffusion measures $\mathbf{f(r)}$ lives in a 3-dimensional space. Computing the joint PDF of the 3-dimensional space is computationally intensive. Further, the measures N, FA, M_d are mutually orthogonal. As such, we compute a 1D PDF for each measure separately using (3). Note that, each of these measures captures different aspects of the variation in "shape" of the diffusion tensor and thus are independent of the orientation.

Several schizophrenia studies [15] have shown abnormalities in the white matter region of the brain. We thus choose this entire region (white matter) to compute the PDF. Specifically, a diffusion tensor is estimated at each voxel and FA is computed in the entire image volume. Regions of the brain that have $FA \geq 0.4$ are selected for further analysis (see Fig. 2). This roughly corresponds to the single fiber white matter region in the brain. Note that, we chose a threshold of 0.4 in order to exclude regions that have crossing fiber bundles (which would result in lower FA). Such crossing regions cannot be correctly represented using a single diffusion tensor. All the other features (such as, M_d, N) are computed in this region (with $FA > 0.4$).

We should note that since FA is a discriminatory feature between the two populations (first-episode schizophrenics and healthy controls), thresholding the image in itself amounts to a feature selection step. For example, if one group in

general has lower FA than another, this would lead to a difference in the estimated pdf which would be useful during classification.

Using (3), we compute the PDF for each of the three discriminatory measures and combine them into a matrix representation denoted by $\mathbf{p} = [p_n \;\; p_{fa} \;\; p_{md}]$. Thus, each patient scan i can now been transformed into a probabilistic representation (matrix) $\mathbf{p}^i$ of dimension $n_b \times 3$, where n_b is the number of bins used in the pdf computation. In our subsequent discussions, we will use this representation in our classifier.

Figure 3a–c show the PDF's for 22 first-episode (FE) schizophrenic patients (red) along with 20 age-matched normal controls (NC) (blue). A visual inspection shows differences between the two groups (blue and red) for each of these measures. Figure 4 shows a cross-section of the two groups for a certain value of FA. This figure confirms the existence of two distinct clusters in the data (albeit with overlap).

2.3 *Parzen Window Classifier*

The Parzen window classifier was first introduced by Jain and Ramaswami [11]. In this method, a Parzen window based density estimate [9] is used to compute the likelihood that a new data point belongs to one of the groups in the training data set.

Let $\{\mathbf{p}^i_{fe}\}_{i=1}^{N_{fe}}$ and $\{\mathbf{p}^i_{nc}\}_{i=1}^{N_{nc}}$ be the set of N_{fe} FE and N_{nc} NC subjects in the training data set. Given a test data point $\hat{\mathbf{p}}$, the likelihood (probability) that it belongs to either group can be computed using the Parzen window density estimator as follows:

$$\begin{aligned} \mathscr{P}^{fe}(\hat{\mathbf{p}}) &= \frac{1}{N_{fe}} \sum_{i=1}^{N_{fe}} K(\hat{\mathbf{p}}, \mathbf{p}^i_{fe}), \textit{aligned} \\ \mathscr{P}^{nc}(\hat{\mathbf{p}}) &= \frac{1}{N_{nc}} \sum_{i=1}^{N_{nc}} K(\hat{\mathbf{p}}, \mathbf{p}^i_{nc}), \end{aligned} \tag{4}$$

where $K(.,.)$ is a Gaussian kernel given by

$$K(\mathbf{p}^i, \mathbf{p}^j) = \exp\left(\sum_{m=1}^{3} \frac{- \parallel p^i_m - p^j_m \parallel^2}{\sigma^2_m} \right), \tag{5}$$

with $m = [N, FA, M_d]^T$ as described earlier, and i, j represent the indices for ith and jth subject. Note that, we assume that the PDF's of each of the diffusion measures for a subject are independent, due to the fact that these measures themselves are orthogonal.

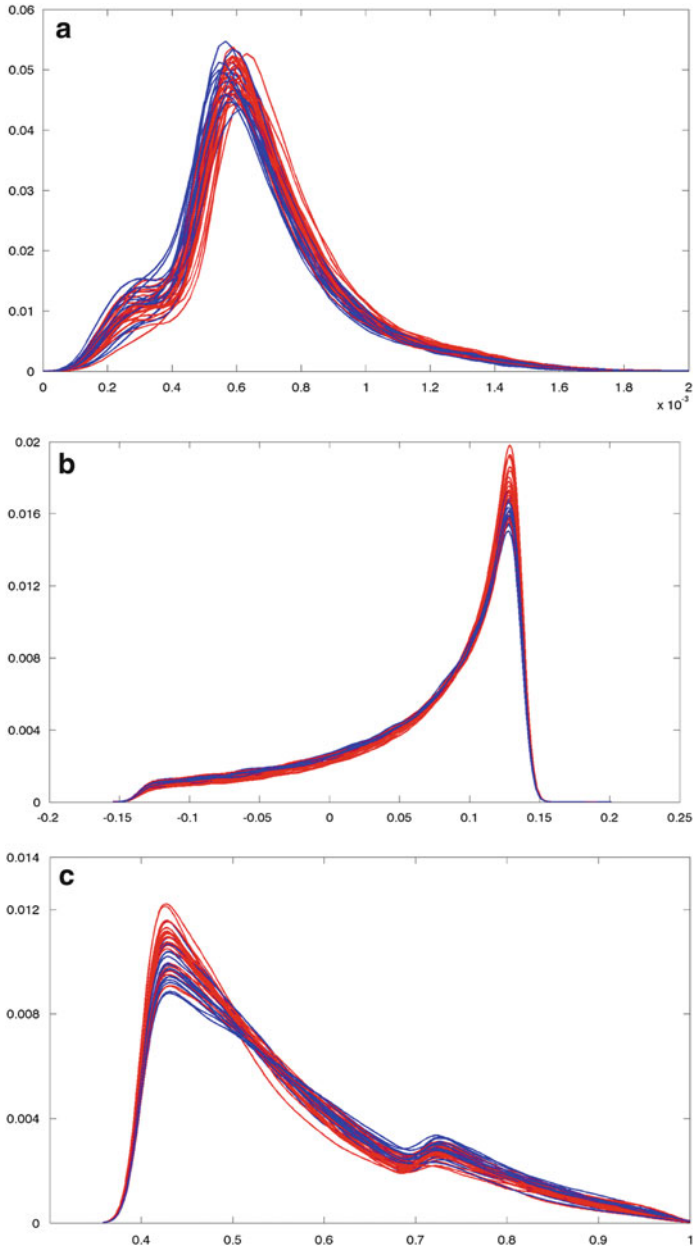


Fig. 3 Probability density functions of various anisotropy measures for 22 FE patients (*red*) and 20 NC (*blue*). (**a**) Norm. (**b**) Mode. (**c**) FA

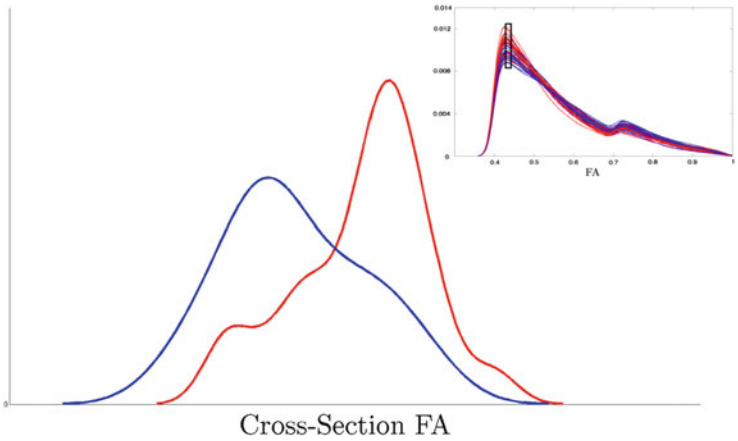


Fig. 4 Cross-sectional distribution of the PDF's of FA (*upper right*) for FE (*red*) and NC (*blue*) subjects

2.3.1 Design Choices

For each of the two groups, we choose σ_m using the following relation:

$$\sigma_m = \frac{c_m}{N} \sum_{i=1}^{N} \min_{i \neq j} \parallel p_m^i - p_m^j \parallel, \quad j = 1, 2, ..N, \quad m = 1 \text{ to } 3,$$

where N is N_{fe} for the group of FE patients and $N = N_{nc}$ for NC subjects. Thus a different set of $\{\sigma_m\}_{m=1}^3$ is computed separately for each group in the training data set. The constant c_m is a scalar that is computed so that the training error is minimized. Typical values for c_m that give a good generalization of the sampled data while reducing the risk of over fitting lie in the range $c_m \in [1.5,\ 2]$, as has been noted in [7]. In numerical experiments, we discretize c_m in the range $[1.5,\ 2]$ at an interval of 0.1. The value of c_m that minimizes the training error is chosen for a given training data set. We should note that this is the only parameter one needs to choose in our entire classification system.

This data driven approach of choosing σ_m is quite common in the literature and has been used in other works as well [7]. This choice of σ_m is guided by the following considerations: (1) σ_m varies appropriately with the scaling of each of the components of m, (2) It minimizes the training error of the classifier, (3) it respects the distribution of points within the clusters (whether the points are spread out or densely packed).

Thus, from the probabilities obtained in (4), we obtain the following simple classification rule:

$$\text{Classification of } \hat{\mathbf{p}} = \begin{cases} \mathscr{P}^{fe}(\hat{\mathbf{p}}) > \mathscr{P}^{nc}(\hat{\mathbf{p}}), & \hat{\mathbf{p}} \text{ is FE patient} \\ \mathscr{P}^{fe}(\hat{\mathbf{p}}) \leq \mathscr{P}^{nc}(\hat{\mathbf{p}}), & \hat{\mathbf{p}} \text{ is NC.} \end{cases}$$

3 Results

3.1 Data Acquisition Protocol

Our dataset consisted of 22 FE patients with average age 20.89 ± 4.8 years and 20 NC with average age 22.3 ± 4.2 years ($p = 0.21$). All the subjects were scanned as part of Dr. Martha Shenton's NIH grant (R01 MH 50740) on a 3-T GE system using an echo planar imaging (EPI) diffusion weighted image sequence. A double echo option was used to reduce eddy-current related distortions. To reduce impact of EPI spatial distortion, an eight channel coil was used to perform parallel imaging using Array Spatial Sensitivity Encoding Techniques (GE) with a SENSE-factor (speed-up) of 2. Acquisitions have 51 gradient directions with b-value $= 900$ and eight baseline scans with $b = 0$. The original GE sequence was modified to increase spatial resolution, and to further minimize image artifacts. The following scan parameters were used: TR 17,000 ms, TE 78 ms, FOV 24 cm, 144×144 encoding steps, 1.7 mm slice thickness. All scans had 85 axial slices parallel to the AC-PC line covering the whole brain.

The raw diffusion weighted images were preprocessed using the Rician noise removal algorithm of [2] followed by eddy current and head motion correction algorithm [13] (part of the FSL package – http://www.fmrib.ox.ac.uk/fsl/flirt/).

4 Classification Results

4.1 Leave-One-Out Cross-Validation

Leave-one-out (LOO) is an unbiased technique for cross-validation of classification results especially when the training data set is small [5, 23]. This is one of the techniques we use to test our classifier. In this method, one subject is removed from the dataset and the classifier is trained on the remaining samples. This procedure is repeated for all available samples and classification results are computed.

In our case, the data samples are the matrices $\mathbf{p}^i$ of dimension $(n_b \times 3)$, with each column representing a discretized pdf of the feature vectors. Here, n_b is the number of bins, which we fix to 300 in all experiments. Given the matrices $\mathbf{p}^i$ for all subjects, the probability of a previously unseen subject is computed using Eq. (4). This procedure is repeated by removing one datum each time and using

Table 1 95 % confidence intervals on the performance of the classifier for LOO experiment

Method	Lower limit	Upper limit	Best estimate
Bayesian	0.788	0.968	0.901
Adjusted Wald	0.768	0.967	0.883
Exact Wald	0.811	0.993	0.892

Table 2 Classifier performance for LOO experiment with different features

Detection rate	False positives	Features used
0.772	0.350	FA
0.909	0.250	FA, MD
0.909	0.100	FA, MD, N

the remaining samples as training data set. Thus, one sample is used as test while the remaining samples are used to train the classifier. The correct detection rate is then computed by counting the number of times the test sample was correctly identified (FE or NC) while testing all the subjects (in our case it is 42). The false positive rate is given by the number of subjects that were "predicted" by the classifier as FE, whereas they were NC. The overall classification error is given by the number incorrect classifications "predicted" by the classifier. In our experiments, **the detection rate (true positives) obtained for LOO cross-validation is** 90.91 %**, while the false positive rate is** 10.0 %**. The overall classification error is** 9.52 %**.**

As has been pointed out by the authors in [10], for small sample size, it is not enough to validate the results using LOO experiment. Instead, one should compute confidence intervals that give a lower and upper bound on the performance of the classifier. Several methods have been proposed in the literature to compute these bounds for small sample size, of which the Bayesian and Binomial bounds are most popular.

Table 1 gives the 95 % Bayesian and Binomial (Exact Wald and Adjusted Wald) [17, 21] confidence intervals (upper and lower limit) on the overall performance of the classifier. Intuitively, a 95 % confidence interval indicates that in 95 out of 100 experiments, the overall performance of the classifier will fall within the stated confidence interval. These confidence intervals are also a function of the number of samples in the data set. Thus, as the number of samples tested increases, the confidence interval becomes narrow and converges to the "true" estimate [10, 12]. The Exact method was designed to guarantee at least 95 % coverage, whereas the approximate methods (adjusted Wald) provide an average coverage of 95 % only when a large number of samples are available.

The above LOO experiment included all the three components of vector f as features. Table 2 shows classification results for LOO experiment, but with different number of features. As is clear, including all the three features does improve the performance of the classifier. Adding more features such as radial diffusivity, linear anisotropy, etc. did not improve the performance of the classifier.

5 Discussion

In this paper, we proposed a novel probabilistic classification method for separating first-episode schizophrenic patients from age-matched normal controls using anisotropic measures derived from diffusion tensor images. The output of the classifier is a probabilistic score of a previously unseen subject being FE or NC. We validate the proposed classifier using a leave-one-out experiment obtaining a sensitivity of 90.91 % and specificity of 90 %. In this work, we chose the entire white matter to perform classification. However, individual fiber tracts such as corpus callosum, fornix, cingulum bundle, etc. may be able to provide more information regarding the variation of these fiber bundles in either population. Our future work entails examining these fiber tracts to detect abnormalities and subsequently use them for probabilistic classification. We should note that the methodology presented here is quite general and can be applied for classification of many other types of brain disorders (bipolar disorder, schizotypal personality disorder, etc.).

This work is a first step towards early detection of schizophrenia, which can result in better patient care. Further, the probabilistic methodology proposed in this work could be used to study the effect of medication by analyzing changes in white matter anisotropy.

Acknowledgements This work has been supported in part by a Department of Veteran Affairs Merit Award (Dr. Martha Shenton), the VA Schizophrenia Center Grant (MS), NIH grant R01MH50740 (MS), R01MH097979 (Rathi) and R01 MH074794(Westin).

References

1. Aitken, A.: On least squares and linear combination of observations. In: Proceedings of the Royal Society of Edinburgh, vol. 55, pp. 42–48 (1934)
2. Aja-Fernandez, S., Niethammer, M., Kubicki, M., Shenton, M.E., Westin, C.F.: Restoration of DWI data using a rician LMMSE estimator. IEEE Trans. Med. Imaging **27**, 1389–1403 (2008)
3. Basser, P., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**(1), 259–267 (1994)
4. Caan, M., Vermeer, K., van Vliet, L., Majoie, C., Peters, B., den Heeten, G., Vos, F.: Shaving diffusion tensor images in discriminant analysis: a study into schizophrenia. Med. Image Anal. **10**(6), 841–849 (2006)
5. Cawley, G., Talbot, N.: Efficient leave-one-out cross-validation of kernel Fisher discriminant classifiers. Pattern Recognit. **36**(11), 2585–2592 (2003)
6. Chenevert, T., Brunberg, J., Pipe, J.: Anisotropic diffusion in human white matter: demonstration with MR techniques in vivo. Radiology **177**(2), 401–405 (1990)
7. Cremers, D., Kohlberger, T., Schnrr, C.: Nonlinear shape statistics in mumford-shah based segmentation. In: 7th ECCV '02, Copenhagen, vol. 2351, pp. 93–108, (2002)
8. Davatzikos, C., Shen, D., Gur, R., Wu, X., Liu, D., Fan, Y., Hughett, P., Turetsky, B., Gur, R.: Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. Arch. Gen. Psychiatry **62**(11), 1218–1227 (2005)
9. Girolami, M.: Orthogonal series density estimation and the kernel eigenvalue problem. Neural Comput. **14**(3), 669–688 (2002)

10. Isaksson, A., Wallman, M., Goransson, H., Gustafsson, M.: Cross-validation and bootstrapping are unreliable in small sample classification. Pattern Recognit. Lett. **29**(14), 1960–1965 (2008)
11. Jain, A., Ramaswami, M.: Classifier design with Parzen windows. In: Pattern Recognition and Artificial Intelligence, North Holland, pp. 211–228 (1988)
12. Jaynes, E.: Confidence intervals vs. Bayesian intervals. Found. Probab. Theory Stat. Inference Stat. Theor. Sci. **2**, 175–257 (1976)
13. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage **17**(2), 825–841 (2002)
14. Kindlmann, G., Ennis, D., Whitaker, R., Westin, C.: Diffusion tensor analysis with invariant gradients and rotation tangents. TMI **26**(11), 1483–1499 (2007)
15. Kubicki, M., McCarley, R., Westin, C.-F., Park, H.-J., Maier, S., Kikinis, R., Jolesz, F., Shenton, M.: A review of diffusion tensor imaging studies in schizophrenia. J. Psychiatr. Res. **41**, 15–30 (2007)
16. LeBihan, D., Mangin, J., Poupon, C., Clark, C., Pappata, S., Molko, N., Chabriat, H.: Diffusion tensor imaging: concepts and applications. J. Magn. Reson. Imaging **13**, 534–546 (2001)
17. Lewis, J., Sauro, J.: When 100% really Isn't 100%: improving the accuracy of small-sample estimates of completion rates. J. Usability Stud. **1**(3), 136–150 (2006)
18. Ozarslan, E., Vemuri, B., Mareci, T.: Generalized scalar measures for diffusion MRI using trace, variance, and entropy. Magn. Reson. Med. **53**(4), 866–876 (2005)
19. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Stat. **33**(3), 1065–1076 (1962)
20. Pohl, K.M., Sabuncu, M.R.: A unified framework for MR based disease classification. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) Information Processing in Medical Imaging. Lecture Notes in Computer Science, vol. 5636, pp. 300–313. Springer (2009). ISBN 978-3-642-02497-9
21. Sauro, J., Lewis, J.: Estimating completion rates from small samples using binomial confidence intervals: comparisons and recommendations. In: Human Factors & Ergonomics Society (ed.) Proceedings of the Human Factors and Ergonomics Society: 49th Annual Meeting, Orlando (2005)
22. Shenton, M., Dickey, C., Frumin, M., McCarley, R.: A review of MRI findings in schizophrenia. Schizophr. Res. **49**(1–2), 1–52 (2001)
23. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (2000)
24. Westin, C.-F., Maier, S.E., Mamata, H., Nabavi, A., Jolesz, F.A., Kikinis, R.: Processing and visualization of diffusion tensor MRI. Med. Image Anal. **6**(2), 93–108 (2002)

A System for Combined Visualization of EEG and Diffusion Tensor Imaging Tractography Data

Alexander Wiebel, Cornelius Müller, Christoph Garth, and Thomas R. Knösche

Abstract In this paper we present an interactive system that integrates the visual analysis of nerve fiber pathway approximations from diffusion tensor imaging (DTI) with electroencephalography (EEG) data. The technique uses source reconstructions from EEG data to define certain regions of interest in the brain. These regions, in turn, are used to selectively display subsets of the approximated fiber pathways in the brain. The selected pathways highlight potential connections from activated areas to other parts of the brain and can thus help to understand networks on which most higher brain function relies. Users can explore the neuronal network and activity by navigating in an EEG curve view. The navigation is supported by optional mechanisms like snapping to time points with present reconstructed dipoles and visual cues highlighting such points. To the best of our knowledge, the presented combination of time navigation in EEG curves together with DTI pathway selection at the corresponding dipole positions is new and has not been described before. The presented methods are freely available in an open source system for visualization and analysis in neuroscience.

Keywords Diffusion tensor imaging • Tractography • Visualization • EEG • Reconstructed Dipoles • GUI

A. Wiebel (✉)
Zuse Institute Berlin (ZIB), Berlin, Germany

Coburg University of Applied Sciences, Coburg, Germany
e-mail: alexander.wiebel@hs-coburg.de

C. Müller • C. Garth
Technische Universität Kaiserslautern, Kaiserslautern, Germany

T.R. Knösche
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2_15,

1 Introduction

Neuroscience and neurology employ a large variety of measurement techniques to examine anatomy, function and pathologies of the human brain in vivo. Among them are electroencephalography (EEG), diffusion weighted magnetic resonance imaging (dwMRI), magnetoencephalography (MEG), and computed tomography (CT). Given the very complex structure of the brain, combining these techniques is crucial for gaining deeper insight into brain functions or diseases. In other words, multi-modal visualization and analysis tools are needed to provide neuroscientists and physicians the support they need for advanced research or diagnosis.

The work described in this paper is motivated by findings [28] that show correlations between epileptical networks measured using MEG or EEG, and connectivity of white matter fiber tracts. We present a system that enables the simultaneous inspection of EEG measurements and diffusion tensor imaging (DTI) data obtained from dwMRI measurements. In particular, a combination of dipole locations reconstructed from EEG (see e.g. [13]) and fiber pathways (see e.g. [2, 10, 16, 18], Fig. 1) obtained from the DTI data is used for visualization, and an intuitive navigation for exploring the time dimension of the data is provided on the basis of simple curve plots of the measured EEG data. A user navigating in time triggers a search for the dipoles active at the current time. The positions of the found dipoles are then used to select fibers that run through activated brain areas represented by the dipoles (Fig. 2).

This way not only the brain activation related to a particular experimental or clinical condition can be visualized, but also the potential connections from the activated areas to other parts of the brain, i.e. the potentially activated networks. This is important as most higher brain function is thought to rely on widespread networks. Similar results can be achieved by a combination of fMRI and DTI [8, 15, 21]. However, in contrast to fMRI, EEG can achieve a very high resolution in time and thus is able to capture very fast current changes in the brain, which can occur within the order of milliseconds.

An illustration providing an overview of the overall pipeline of our approach is given in Fig. 3. It consists of the measurement, a preprocessing step and an *interactive visualization and analysis loop*. Throughout the paper we first describe how our approach is related to previous work (Sect. 1.1), then we provide a brief introduction to the preprocessing (Sect. 2) and finally introduce the new technique (Sect. 3) as well as implementation details (Sect. 4). The actually novel technique is integrated in the *interactive visualization and analysis loop*.

1.1 Related Work

The literature dealing with advanced EEG visualization is relatively sparse. Advanced in this context means showing more than the measured voltage curves

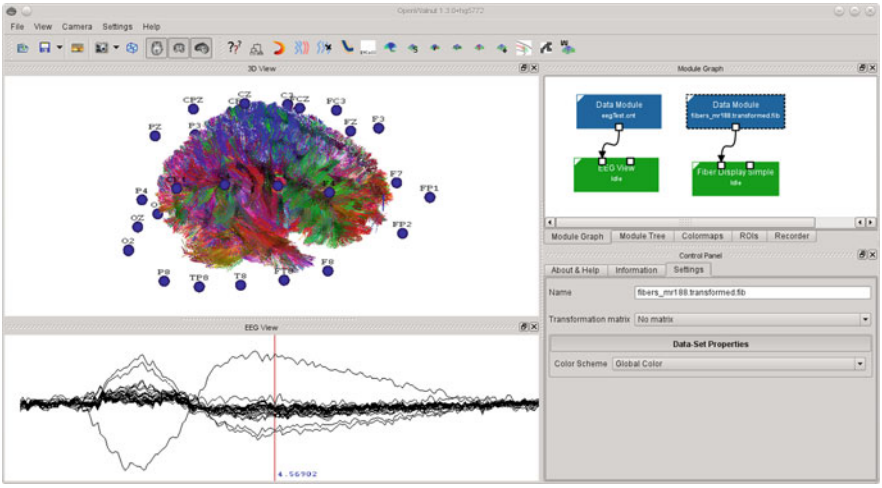


Fig. 1 DTI fibers tracts rendered in context of electrode positions in upper part of window and butterfly plot (details in Sect. 3.2) of EEG recording in lower part

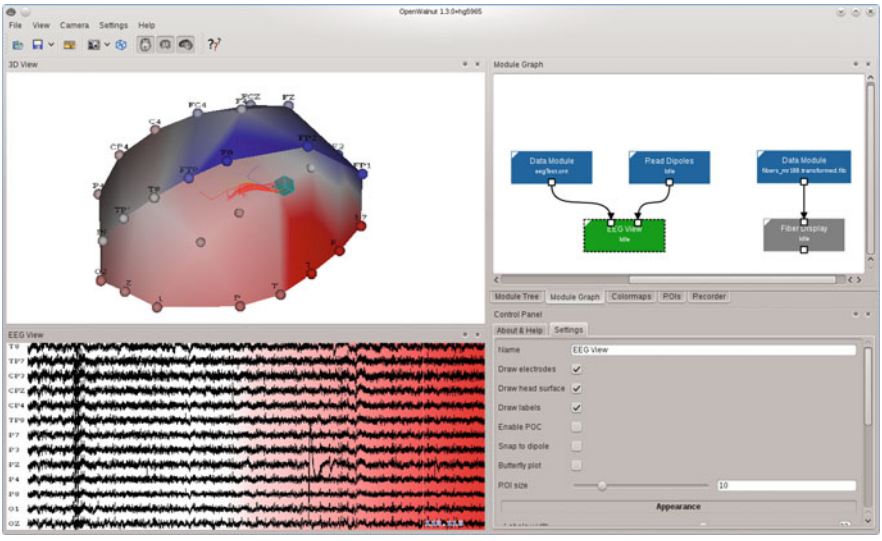


Fig. 2 DTI fibers tracts selected by the position of the reconstructed dipole ($10 \times 10 \times 10$ mm ROI box) are rendered in context of head surface with voltage at the different electrodes mapped to color in the upper part of window. The lower part shows a common curve display of the EEG recording ("EEG View"). The "EEG View" widget is used for navigation in time (notice the vertical line marker, $t = 118.113$). The color coding in the background indicates the presence of a loaded dipole. The color represents the magnitude of the dipole

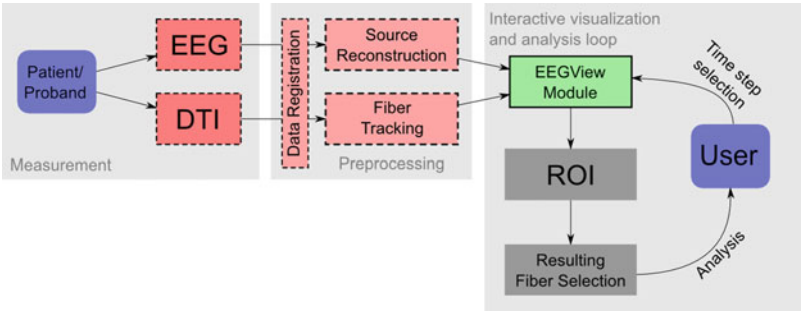


Fig. 3 Pipeline illustrating the overall approach. The central part of the system is integrated in the *interactive visualization and analysis loop*

(Fig. 2), a butterfly plot (Fig. 1) or frequency plots of the data. Exceptions are work on visualizing EEG recordings in 3D by Sourina et al. [23] and an information visualization approach by ten Caat which he summarized in his doctoral thesis [25]. One of ten Caat's results is a method to visualize the coherence in EEG data as a network of interconnected functional units.

Some work regarding the analysis and visualization of sources reconstructed from EEG and MEG data (see Johnson et al. [11] and Wolters et al. [30]) has been performed. Wolters et al. [30] analyze result of source reconstructions from EEG data and their connection to data obtained from dwMRI. However, they do not employ the diffusion weighted data for approximating white matter fiber pathways for visualization, but to obtain conductivity tensors. The tensors are used to analyze the influence of the white matter conductivity on EEG/MEG field computation.

Discussions concerning the relation between functional MRI (fMRI) and EEG source reconstructions are quite common in the literature (see e.g. [27]). Also, the combination of fMRI and DTI-based fiber tracking to define and analyze white matter pathways is well established [8, 15, 21]. Such a combined analysis can be performed in many widely used software packages [9, 26]. In contrast, the only work we found, that visualizes EEG data and fiber tracts, is the description of the NUTMEG system by Dalal et al. [4]. Their system, in conjunction with Xipy (also mentioned in the paper), shows white matter fiber tracts in conjunction with visualizations of functional maps obtained from source reconstruction. The system does not seem to provide an interactive navigation mechanism for inspecting reconstructions at different points in time and the original EEG recording curves are not visible to the user. It is furthermore unclear how the fiber tracts are selected.

In summary, there has been much work combining MRI and EEG (see also Tadel et al. [24]), few work dealing with DTI-based pathways together with EEG and no work allowing time navigation based on the display of voltage curves. This paper and the system presented in it aim at filling the gap left in the latter two areas.

2 Preprocessing

Acquiring the data during EEG and MRI measurements is only the very first step in a pipeline leading to the successful analysis of the subject's condition. The data needed for the effective utilization of the technique presented in this paper, have to be registered, tractography has to be applied to the DTI and dipoles have to be reconstructed from the EEG. As illustrated in Fig. 3, we consider these three parts of the pipeline as preprocessing steps and will detail two of them (tractography and reconstruction) in the following because the combination of their results lies in the heart of the presented technique. Naturally, the registration has to be as good as possible, but this is the case for most multi-modal visualization and analysis approaches and will thus not be discussed in the current context.

2.1 DTI Tractography

Approximated fiber pathways are one basis for the visualization and analysis of our approach. The data used for the fiber approximation in our example comes from a DTI measurement acquired with a 3-T scanner *Siemens 3T Trio* at the Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. The DTI image is given as second order tensors at a resolution of 1.72 mm. We used the *tensorlines* method introduced by Weinstein et al. [12, 29] as implemented in MedINRIA [6] for generating all tracts used in this paper. To obtain a full brain tractogram, the tensorlines were seeded at voxels where the fractional anisotropy exceeds a certain threshold. However, other seeding strategies are possible as the used strategy does not imply any constraints on the design and implementation of the parts of the *interactive visualization and analysis loop*. The reconstructed tracts are stored into a file from which they can be loaded for the following steps as needed. The complete tractogram in our example consists of 74,313 polylines represented by 5,397,993 vertices.

2.2 EEG Source Reconstruction

Electroencephalography (EEG) is a technique measuring the accumulated electrical activity inside the brain by recording voltage changes at the head surface. Compared to other brain measurement techniques EEG is recorded at relatively few (21–256) locations [19]. Although EEG is recorded from electrodes placed *around* the head surface and although the properties (amplitudes, frequencies, etc.) of the recorded voltage curves are expressive and meaningful themselves, one is interested in local currents *inside* the head because they are indicative for neuronal activity at the different locations in the brain. The methodology applied to obtain these currents is called *source reconstruction*. Considering, that the density of the source current

at a certain location is a vector with three components, the problem of source reconstruction for more than 42 locations from, e.g., 128 electrodes, is underdetermined. This fact is a first hint at the comparatively low spatial accuracy of reconstructed sources. To get a higher accuracy, further constraints obtained from prior knowledge of the brain structure have to be imposed [19]. Still, the achievable spatial resolution is much worse than that of MRI or CT. Despite these disadvantages, EEG is an important tool for examining brain activity and function because of its high resolution in time. Together with the often simultaneously discussed magnetoencephalography (MEG), EEG is the only common measurement technique that is able to capture the fast current changes in the brain, which can occur within milliseconds.

In this paper, we use the ASA software package [31] for source reconstruction and dipole localization. ASA (Advanced Source Analysis) is a highly flexible package for analysis of continuous and event-related EEG/MEG signals. Regarding our task, it provides a variety of useful source reconstruction and signal analysis features. For single dipoles a spatio-temporal dipole fit as described by Scherg [19] is used. A tutorial for using ASA for source reconstruction can be found here: [1].

Like the tracts, the EEG and the reconstruction results are stored into files from which they can be loaded for the following steps as needed.

3 Technique

After preprocessing has been performed, the data are ready for exploration using the technique proposed in this paper (see Fig. 3). Our specific approach is a combination of visualization and interaction techniques. In the following subsection, we first describe the visualization of the fibers and of the EEG on the head surface. Afterwards, we describe how we present the original EEG signal and how this presentation is used for interactively navigating through the time-dependent data.

3.1 Visualization

The fibers are visualized using the techniques available in OpenWalnut [5]. These include simple lines (Fig. 2), illuminated lines (Fig. 1), and fake tubes (Fig. 4). Each of these line types can be rendered with a user-defined constant color as well as with either local or global directional color coding.

Rendering only a subset of a set of fibers, e.g. of a whole brain tractogram, can be achieved by defining regions of interest (ROIs). As a result only those fibers which run through the defined ROIs will be shown. This selection mechanism is an implementation of the approach reported by Blaas et al. [3]. As the selection mechanism allows the combination of different ROIs for selecting a single fiber bundle, it is also possible to select those fibers that run through the locations of several simultaneously active dipoles. Furthermore, multiple bundles each running through one of the ROIs can be selected. An example of the latter is shown in Fig. 4.

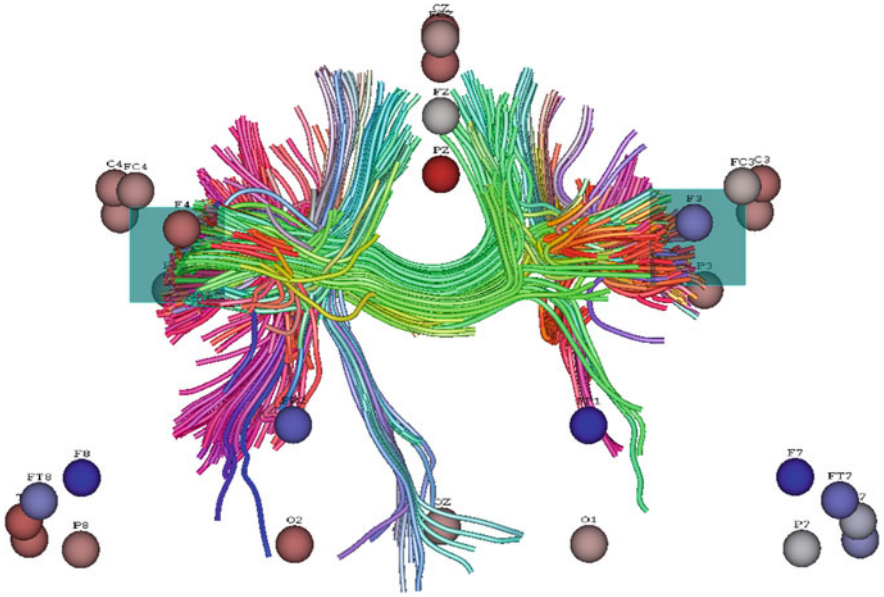


Fig. 4 Two ROI boxes (20 mm) selecting fibers are the result of two simultaneously active dipoles (t = 128.433)

The presentation of the original EEG signal is described in Sect. 3.2 because this presentation is also used for navigation in time.

In order to give a quick overview of the magnitude distribution at the currently selected point in time, we color-code the magnitudes of the different channels on the head surface (see Fig. 2). In our case the "head surface" is a simple interpolation of the electrode positions (Delaunay triangulation). We use a 2D implementation of the triangulation algorithm. Therefore we have to project the positions to two dimensional space first. A projection along the axial direction yields satisfactory results most of the time. In some cases, however, where the electrode positions cover more than the upper hemisphere, i.e. where parts of the projection overlap, we have to *spread* the lower positions somewhat before performing the projection. This prevents overlapping triangles in the projection and thus guarantees a valid triangulation. To color the triangles connecting the electrode positions, we map values, that where linearly interpolated from the values at the electrodes, to color.

3.2 Navigation

The user interface for navigating through the EEG data along the time dimension is visible in the lower left widget in Fig. 2. It is realized as a kind of timeline widget [20]. As in most EEG software, the standard presentation shows the signals of the different electrodes (channels) as curves in row-oriented layout. The name

of the electrode belonging to a certain row is always visible on the left. A vertical red line indicates the currently selected point in time. The position of this line and thus the selected time value can be adjusted by left-clicking in the widget. This very simple type of interaction makes data exploration in time feasible and intuitive for all (also new) users. As exploring the correspondence of the instantaneous EEG signal and the fibers selected for dipoles at the same instant of time are the main goal of this work, the simple time point selection is central to the interaction and navigation. Dragging with the middle mouse button allows to bring additional channels (moving up/down) or other time intervals (moving left/right) of the recording into view. This is important for exploring the signal captured at all electrodes and thus to get an impression of the complete momentary state of the EEG. Dragging (left/right) with the right mouse button zooms (out/in) the currently visible time interval. The mouse wheel adjusts the scaling of the curves, i.e. the mapping from the measured voltage to the height of the curve. The different zoom and scaling levels are useful for getting an overview or inspecting details of the EEG curves. All mentioned adjustable values (position, scaling, etc.) can also be adapted in the control panel (right in Fig. 2). This allows to select specific numeric values by typing the exact numbers and can be useful for quantitative studies where the exact values are needed.

An alternative EEG representation, called butterfly plot, which superimposes the signals of all channels, is visible in Fig. 1. This representation highlights certain particularly interesting instants of time, while, due to visual clutter, the single curves are not discernible anymore. All interaction types described for the conventional representation are also available for the butterfly plot. As a result, navigation in time is possible in both representations by simply clicking at the position of the desired instant of time.

If immediate source reconstruction in the background was possible, we could simply determine active locations and thus ROIs (see Sect. 3.3) for any selected instant of time. However, as source reconstruction is a time consuming process and as we have to rely on externally performed dipole localization, we only have dipoles for a number of time intervals. These time intervals are highlighted in the EEG plot window as shown in Fig. 2. A white background means no dipole; color indicates the presence and magnitude (by color saturation) of a loaded dipole.

When the user selects a certain point in time, the system checks if there are active dipoles at this time or in other words, if the selected instant of time lies in a time interval of one of the dipoles. The positions of the respective dipoles (one or multiple) are then used for fiber selection. The selection will be described in the Sect. 3.3 in more detail.

Scrolling through the whole EEG plot, searching for dipoles can be a tedious task for plots representing a long recording session. Therefore, we introduced another navigation mode, which we call *snap to dipole* (see Fig. 2, "Settings" tab). This mode is tailored to support fast selection of times for which dipoles have been loaded. If this mode is active and the user clicks, i.e. selects a certain point in time, the time marker *snaps* to the next position in time where a dipole can be found. One can think of this as a nearest neighbor search. The position located by the snapping will

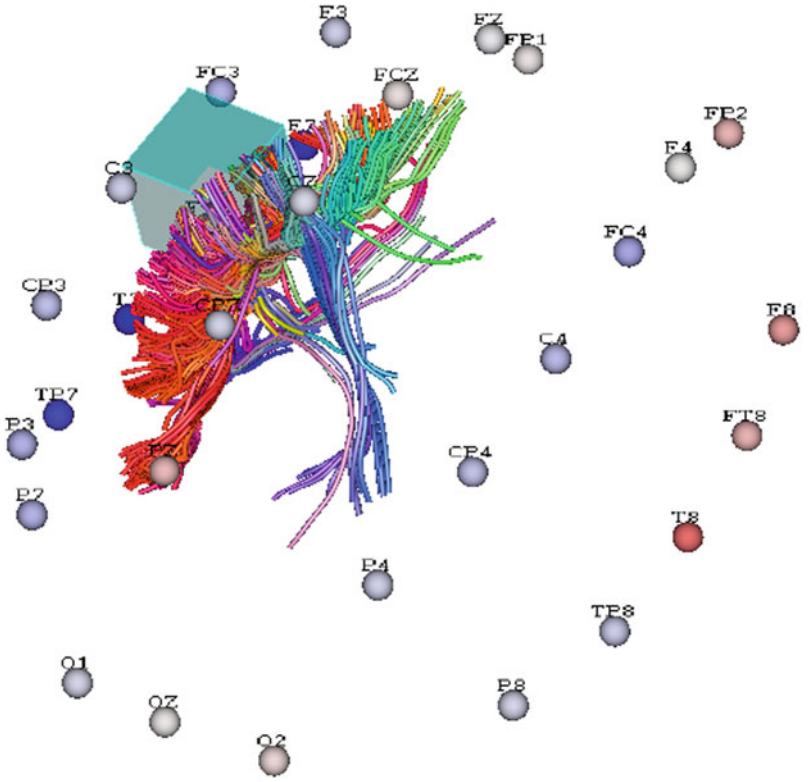


Fig. 5 Fiber selection resulting from ROI box with 3 mm edge length located at reconstructed dipole position (t = 117.889)

be the border of an interval containing an active dipole. The dipole magnitude will thus be relatively low at this point. With the color coding, however, it is now easy to navigate to positions at which the magnitude is higher. The snapping is especially important because the active dipoles often cover only a small portion of the whole time of the loaded EEG data.

3.3 Combination

As alluded to before, the combined visualization or combined exploration of fiber and EEG data is achieved by generating regions of interest (ROIs), in our case boxes, representing the dipoles active at the currently selected point in time.

The generated ROI box is centered at the given dipole position. As a first approximation we use a box with $30 \times 30 \times 30$ mm extent. A selection exemplifying the result of such a ROI is shown in Fig. 5. The initial box size is a conservative

choice in the sense that it selects relatively many fibers instead of showing too few and thus missing the important ones. See Scherg [19] for typical localization errors that led to our choice of the box size. The user can decrease the size of the ROI to explore fibers that would be of interest if the activation was smaller or more accurately localizable than the initial region. When changing its size, the ROI will grow or shrink equally in all directions and will thus stay centered at the dipole position.

As described in Sect. 3.1, the fiber selection mechanism is very flexible in combining the effects of different ROIs. We employ this for dealing with the case of multiple simultaneously active dipoles. The usual use case is to show all fibers that are related to any of the dipoles. This is demonstrated in Fig. 4 for two dipoles. A second option is to select only those fibers that run through all ROIs at the same time. However, one should be aware that this can results in showing no or only very few fibers for some dipole configurations. Even in the examples shown in Fig. 4, there is no fiber that runs through both ROIs. Actually, only one fiber of the corpus callosum is selected by the right ROI and this fiber does not pass through the ROI in the left part of the image. Figure 5 shows that even increasing the ROI size results in the selection of only two fibers of the corpus callosum.

4 Implementation Details

Most data stemming from measuring, e.g. imaging, the brain is small enough to easily fit into the main memory of current commodity hardware. EEG data, in contrast, can become very large due to the number of electrodes used and the time recorded. In principal, increasing recording time can make the data arbitrarily large.

Handling such data makes it necessary to keep the treated file open and read only the data that are currently needed into main memory. The direct access to the file is realized by a pager (`WPagerEEG`, Fig. 6) which can be implemented for different file formats. The currently supported format is CNT (**co**n**t**inuous EEG data file). For accessing such files we use the open source library Libeep [22]. As the metadata contained in the CNT file, like information about the measurement channels, is comparatively small, we load all metadata in the beginning. Only the voltage data, i.e. the data measured for each channel, is loaded on demand by the pager. Modules that need access to the voltage data ask the EEG dataset (`WEEG2`) for it by providing a time interval. The dataset class in turn requests the data from the pager. The pager then loads the requested data from the file if it is not yet present in memory. It presents the data to the dataset and removes unused data from memory.

More details of the implementation (Fig. 6) are described in a master's thesis [17] and the source code of our implementation is available in the online repository of OpenWalnut [5].

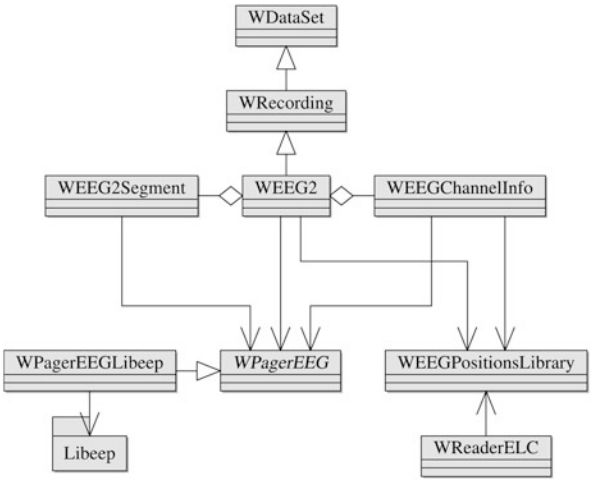


Fig. 6 Data handling scheme

5 Conclusion

The presented system and techniques allow for a new type of visualization and exploration of DTI fiber pathways in conjunction with EEG and MEG source reconstruction data. The combination of the two modalities provides a platform for deeper investigations of the relation between anatomical structures, especially connections, in the brain and certain cognitive functions that these structures are involved in. An intuitive interaction mechanism, specially tailored to the characteristics of the combination of EEG and fiber pathways, has been introduced. This mechanism is the heart of the presented system as it enables the user driven exploratory process.

All presented techniques are freely available in the open source visualization system OpenWalnut (http://www.openwalnut.org). The data can be loaded using the general data import mechanism in combination with the *Read Dipoles* module. The visualization as well as the interaction techniques presented in this paper are integrated in the *EEG View* module.

One of our future research directions will be the relatively straightforward extension of the framework to MEG instead of EEG data. Further possible directions are the combination of the presented fiber bundle selection mechanism with bundle visualization techniques already available in OpenWalnut, e.g. bundle parameter and bundle surface visualization [7], and employing advanced ROI definitions, e.g. on the basis of distributed sources [14], deviating from cubes.

Acknowledgements We thank the OpenWalnut community for providing their software as basis for implementing the presented techniques. We are also grateful to Gerik Scheuermann and his group at the University of Leipzig, who provided an enjoyable and inspiring environment for Cornelius Müller's work on his master's thesis. The reviewers made many very valuable suggestions, we would like to thank them for this. This work was partly supported by the AiF (ZIM grant KF 2034701SS8).

References

1. ANT BV.: ASA Getting Started. ANT BV, Enschede (Sept 2009)
2. Berman, J.I., Chung, S., Mukherjee, P., Hess, C.P., Han, E.T., Henry, R.G.: Probabilistic streamline q-ball tractography using the residual bootstrap. NeuroImage **39**(1), 215–222 (2008)
3. Blaas, J., Botha, C.P., Peters, B., Vos, F.M., Post, F.H.: Fast and reproducible fiber bundle selection in DTI visualization. In: Silva, C., Gröller, E., Rushmeier, H. (eds.) Proceedings of IEEE Visualization 2005, Minneapolis, pp. 59–64 (2005)
4. Dalal, S.S., Zumer, J.M., Guggisberg, A.G., Trumpis, M., Wong, D.D.E., Sekihara, K., Nagarajan, S.S.: MEG/EEG source reconstruction, statistical evaluation, and visualization with NUTMEG. Comput. Intell. Neurosci. **2011**(3), 1–17 (2011)
5. Eichelbaum, S., Hlawitschka, M., Wiebel, A., Scheuermann, G.: OpenWalnut – an open-source visualization system. In: Benger, W., Gerndt, A., Su, S., Schoor, W., Koppitz, M., Kapferer, W., Bischof, H.-P., Pierro, M.D. (eds.) Proceedings of 6th High-End Visualization Workshop, Obergurgl, pp. 76–78 (2010)
6. Fillard, P., Toussaint, N., Pennec, X.: MedINRIA: DT-MRI processing and visualization software. In: Guest Paper at the Similar Tensor Workshop, Las Palmas, Nov 2006
7. Goldau, M., Wiebel, A., Hlawitschka, M., Scheuermann, G., Tittgemeyer, M.: Visualizing DTI parameters on boundary surfaces of white matter fiber bundles. In: Zhang, J. (ed.) Proceedings of the Twelfth IASTED International Conference on Computer Graphics and Imaging, Innsbruck, pp. 53–61. International Association of Science and Technology for Development, ACTA Press, Feb 2011
8. Jang, S.H., Ahn, S.H., Ha, J.S., Lee, S.J., Lee, J., Ahn, Y.H.: Peri-infarct reorganization in a patient with corona radiata infarct: a combined study of functional MRI and diffusion tensor image tractography. Restor. Neurol. Neurosci. **24**(2), 65–68 (2006)
9. Jenkinson, M., Beckmann, C., Behrens, T., Woolrich, M., Smith, S.: FSL. NeuroImage **62**(2), 782–790 (2012)
10. Jeong, W.-K., Fletcher, P.T., Tao, R., Whitaker, R.: Interactive visualization of volumetric white matter connectivity in DT-MRI using a parallel-hardware Hamilton-Jacobi solver. IEEE Trans. Vis. Comput. Graph. **13**(6), 1480–1487 (2007)
11. Johnson, C.R., MacLeod, R., Parker, S.G., Weinstein, D.: Biomedical computing and visualization software environments. Commun. ACM **47**(11), 64–71 (2004)
12. Kindlmann, G.: Visualization and analysis of diffusion tensor fields. PhD thesis, School of Computing, University of Utah (2004)
13. Knösche, T.R.: Solutions of the neuroelectromagnetic inverse problem – an evaluation study. PhD thesis, Universiteit Twente (1997)
14. Knösche, T.R., Praamstra, P., Stegeman, D., Peters, M.: Linear estimation discriminates midline sources and a motor cortex contribution to the readiness potential. Electroencephalogr. Clin. Neurophysiol. **99**(2), 183–190 (1996)
15. Lowe, M.J., Horenstein, C., Hirsch, J.G., Marrie, R.A., Stone, L., Bhattacharyya, P.K., Gass, A., Phillips, M.D.: Functional pathway-defined MRI diffusion measures reveal increased transverse diffusivity of water in multiple sclerosis. NeuroImage **32**(3), 1127–1133 (2006)
16. Mori, S., van Zijl, P.C.M.: Fiber tracking: principles and strategies – a technical review. NMR Biomed. **15**(7–8), 468–480 (2002)
17. Müller, C.: Kombinierte visualisierung von EEG – und diffusions-MRT-Nervenfaser-Daten. master's thesis, Fakultät für Mathematik und Informatik, Universität Leipzig (2010)
18. Reisert, M., Mader, I., Anastasopoulos, C., Weigel, M., Schnell, S., Kiselev, V.: Global fiber reconstruction becomes practical. NeuroImage **54**(2), 955–962 (2011)
19. Scherg, M.: Functional imaging and localization of electromagnetic brain activity. Brain Topogr. **5**(2), 103–111 (1992)
20. Schmidt, V.A.: User interface design patterns. Technical report, DTIC Document (2010)
21. Schonberg, T., Pianka, P., Hendler, T., Pasternak, O., Assaf, Y.: Characterization of displaced white matter by brain tumors using combined DTI and fMRI. NeuroImage **30**(4), 1100–1111 (2006)

22. Smies, R.: Libeep. http://sourceforge.net/projects/libeep/. Last accessed 2012-04-13
23. Sourina, O., Sourin, A., Kulish, V.: EEG data driven animation and its application. In: Gagalowicz, A., Philips, W. (eds.) Computer Vision/Computer Graphics Collaboration Techniques. Lecture Notes in Computer Science, vol. 5496, pp. 380–388. Springer, Berlin/Heidelberg (2009)
24. Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M.: Brainstorm: a user-friendly application for MEG/EEG analysis. Comput. Intell. Neurosci. **2011**, Article ID 879716 (2011)
25. ten Caat, M.: Multichannel EEG visualization. PhD thesis, Institute of Mathematics and Computing Science, University of Groningen (2008)
26. Toussaint, N., Souplet, J.C., Fillard, P.: MedINRIA: medical image navigation and research tool by INRIA. In: Proceedings of MICCAI'07 Workshop on Interaction in Medical Image Analysis and Visualization, Brisbane (2007)
27. Wagner, M., Fuchs, M.: Integration of functional MRI, structural MRI, EEG, and MEG. Int. J. Bioelectromagn. **3**(1), (2001)
28. Weiland, B., Moran, J., Bowyer, S., Smith, B., Barkley, G., Tepley, N.: DTI fiber tracts connect epileptic networks imaged with MEG coherence analysis. In: 16th International Conference on Biomagnetism, Sapporo, Aug 2008. Japan Biomagnetism and Bioelectromagnetics Society
29. Weinstein, D., Kindlmann, G., Lundberg, E.: Tensorlines: advection-diffusion based propagation through diffusion tensor fields. In: Proceedings of IEEE Visualization '99, Los Alamitos, pp. 249–253 (1999)
30. Wolters, C.H., Anwander, A., Tricoche, X., Lew, S., Johnson, C.R.: Influence of local and remote white matter conductivity anisotropy for a thalamic source on EEG/MEG field and return current computation. Int. J. Bioelectromagn. **7**, 203–206(2005)
31. Zanow, F., Knösche, T.: ASA – advanced source analysis of continuous and event-related EEG/MEG signals. Brain Topogr. **16**(4), 287–290 (2004)

Index

C.-F. Westin et al. (eds.), *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization,
DOI 10.1007/978-3-642-54301-2, © Springer-Verlag Berlin Heidelberg 2014