Susanne Heim

Statistical Diffusion Tensor Imaging: From Data Quality to Fiber Tracking.

Cuvillier Verlag

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Statistical Diffusion Tensor Imaging: From Data Quality to Fiber Tracking.

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik

der Ludwig-Maximilians-Universität München

Susanne Heim

February 19, 2007

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.
1. Aufl. - Göttingen : Cuvillier, 2007
Zugl.: München, Univ., Diss., 2007

978-3-86727-326-8

1. Berichterstatter:	Prof. Dr. Ludwig Fahrmeir
2. Berichterstatter:	Prof. Dr. Gerhard Winkler
Ausw. Berichterstatter:	Prof. Dr. Brian Marx
Tag des Rigorosums:	April 20, 2007

© CUVILLIER VERLAG, Göttingen 2007
Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlages ist es nicht gestattet, das Buch oder Teile daraus auf fotomechanischem Weg (Fotokopie, Mikrokopie) zu vervielfältigen.
1. Auflage, 2007
Gedruckt auf säurefreiem Papier

978-3-86727-326-8

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Abstract

Magnetic resonance diffusion tensor imaging (DTI) allows to infere the ultrastructure of living tissue. In brain mapping, neural fiber trajectories can be identified by exploiting the anisotropy of diffusion processes. Manifold statistical methods may be linked into the comprehensive processing chain that is spanned between DTI raw images and the reliable visualization of fibers. In this work, a space varying coefficients model (SVCM) using penalized B-splines was developed to integrate diffusion tensor estimation, regularization and interpolation into a unified framework. The implementation challenges originating in multiple 3d space varying coefficient surfaces and the large dimensions of realistic datasets were met by incorporating matrix sparsity and efficient model approximation. Superiority of B-spline based SVCM to the standard approach was demonstrable from simulation studies in terms of the precision and accuracy of the individual tensor elements. The integration with a probabilistic fiber tractography algorithm and application on real brain data revealed that the unified approach is at least equivalent to the serial application of voxelwise estimation, smoothing and interpolation. From the error analysis using boxplots and visual inspection the conclusion was drawn that both the standard approach and the B-spline based SVCM may suffer from low local adaptivity. Therefore, wavelet basis functions were employed for filtering diffusion tensor fields. While excellent local smoothing was indeed achieved by combining voxelwise tensor estimation with wavelet filtering, no immediate improvement was gained for fiber tracking. However, the thresholding strategy needs to be refined and the proposed model of an incorporation of wavelets into an SVCM needs to be implemented to finally assess their utility for DTI data processing.

In summary, an SVCM with specific consideration of the demands of human brain DTI data was developed and implemented, eventually representing a unified postprocessing framework. This represents an experimental and statistical platform to further improve the reliability of tractography.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Zusammenfassung

Die Diffusionsbildgebung durch Kernspintomographie erlaubt die ultrastrukturelle Charakterisierung von Gewebe in vivo. Unter Ausnutzung der stark anisotropen Diffusionsprozesse, die vor allem in der weißen Substanz des Gehirns vorherrschen, können Nervenfaserverläufe identifiziert werden (Fasertracking). Statistische Methoden sind hierbei auf allen Verarbeitungsstufen zwischen DTI Rohbildern und der Faservisualisierung notwendig. In der vorliegenden Arbeit wurde ein Modell mit räumlich variierenden Koeffizienten auf Basis von penalisierten B-splines (space varying coefficient model; SVCM) entwickelt, das die gleichzeitige Schätzung, Regularisierung und Interpolation eines Diffusionstensorfeldes leistet. Die Schwierigkeiten bei der Implementierung eines solchen Modells, die mit multiplen 3d variierenden Koeffizientenfeldern und großformatigen realen Datensätzen einhergehen, wurden durch die Verwendung spezieller Speicherformate für dünn besetzte Matrizen und durch eine effiziente Modellapproximation bewältigt. Anhand von Simulationsdaten zeigte sich die höhere Genauigkeit der B-spline basierten Tensorschätzung gegenüber der standardmäßigen voxelweisen Tensorschätzung. Im Weiteren wurde das B-spline basierte SVCM mit der bestehenden Implementierung eines probabilistischen Trackingalgorithmus vereinigt. Bei der Anwendung auf Realdaten konnten etwa gleichwertige Fasermuster erzielt werden wie mit der seriell verfahrenden Standardmethode. Der qualitativen Beurteilung der Tensorschätzung zufolge zeigten beide Verfahren eine unzureichende lokale Adaptivität, so dass versucht wurde, diese über eine Waveletfilterung des Tensorfelds zu optimieren. Durch eine Kombination aus voxelweiser Schätzung und Waveletfilterung wurde erwartungsgemäß für die Simulationsdaten eine deutlich höhere lokale Adaptivität erreicht; die Trackingergebnisse wurden hierbei jedoch eher ungünstig beeinflusst. Mögliche Weiterentwicklungen bestehen in der Optimierung des Thresholdingschritts bei der Waveletfilterung und der Implementierung des vorgeschlagenen Wavelet basierten SVCMs, um das volle Potential von Wavelets für die Analyse von DTI Daten auszuloten.

Zusammengefasst wurde ein Modell mit räumlich variierenden Koeffizienten unter besonderer Berücksichtigung der Dimensionalität und weiterer Eigenschaften von DTI Daten entworfen und damit eine simultan vorgehende Methode zur DTI Datenverarbeitung zur Verfügung gestellt. Unmittelbar daran können weitere statistische Funktionen und Simulationsstudien zur Verbesserung von Fasertracking geknüpft werden.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Contents

Outline **1**

1 Introduction to Magnet Resonance Imaging (MRI) **5**

1.1 Basic principles of MRI 7

1.2 Specifications in Diffusion Tensor Imaging (DTI) 12

2 Noise analysis of DTI data using bootstrap **19**

2.1 Introduction 22

2.2 Methods and materials 23

2.3 Results 28

2.4 Discussion 33

3 Space-Varying Coefficients Model (SVCM) using 3d B-splines **37**

3.1 Introduction 38

3.2 Diffusion tensor imaging 41

3.3 The space-varying coefficient model 45

3.3.1 Multidimensional B-splines 46

3.3.2 Multidimensional smoothing with tensor products 48

3.3.3 Successive smoothing with univariate basis functions 52

3.4 Simulation study 56

3.4.1 Supplement 63

3.5 Application to real data 64

3.6 Discussion 66

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3.7 Explanatory remarks . . . 68
3.7.1 Gibbs artifacts . . . 69
3.7.2 Insufficient edge-preservation . . . 70
3.7.3 Star-shaped artifacts . . . 72
3.8 Implementational issues . . . 73
3.9 The 'svcm'-package . . . 74
3.9.1 Getting started . . . 75
3.9.2 Data format . . . 75
3.9.3 SVCM analysis . . . 76
3.9.4 A note on optimization . . . 77
3.9.5 Resolution increment . . . 80
3.9.6 Possible extensions . . . 82

4 Wavelets for DTI 85

4.1 Introduction to wavelets . . . 85
4.1.1 Univariate wavelet basis . . . 87
4.1.2 Wavelet decomposition and reconstruction . . . 91
4.1.3 Multidimensional wavelets decomposition - a 2d example . . . 100
4.1.4 Thresholding and shrinkage . . . 106
4.1.5 Resolution increment . . . 113
4.2 Denoising the diffusion tensor . . . 114
4.3 Simulation study . . . 117
4.3.1 Simulation model . . . 117
4.3.2 Study design . . . 117
4.3.3 Simulation results . . . 119
4.3.4 Comparison to standard approaches . . . 122
4.4 Real life example . . . 129
4.5 Discussion . . . 131
4.6 Proposal for a wavelet based SVCM . . . 133

5 Fiber tracking on the basis of SVCM estimates 137

5.1 Tractography using linear state space models . . . 139

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

5.2 An example: synthetic data 140
5.3 An example: human data 147
5.4 Discussion 152

6 Conclusion and perspective **155**

Danksagung **159**

Bibliography **161**

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Outline

Diffusion tensor imaging (DTI) is an advanced magnetic resonance technique which has opened a new avenue to assess the relationship between brain function, neuroanatomy and pathology. Apart from monitoring major diseases, such as stroke, multiple sclerosis and Alzheimer's disease on the basis of several quantities describing ultrastructural tissue properties, clinical neuroscience greatly benefits from the possibility to non-invasively map detailed subsystems and circuits interconnecting neuronal centres. The latter application is known as fiber tracking, also referred to as tractography. With its integrative capacity for the build-up of atlases and, in conjuction with functional imaging for the investigation of networking regions, fiber tracking allows for a better understanding of the human brain.

Biophysically, DTI and in particular fiber tracking relies on the strongly anisotropic diffusion of water molecules in cerebral white matter. Using special acquisition schemes, the local process can be assessed by calculating a (3×3)-dimensional diffusion tensor for each volume element (voxel) of the brain. The spatial tensor field provides information on the extent of anisotropy and on local fiber orientation. Importantly, the typical size of an imaging voxel (between $1 \times 1 \times 1$ and $3 \times 3 \times 3\,\mathrm{mm}^3$) largely exceeds the scale of the reflected microscopical diffusion processes on the cellular level. The limited spatial resolution of diffusion images leads to partial volume effects with a mixture of directionalities within one voxel and consequently to ambiguity for fiber tracking algorithms. Various sources of noise inherent to DTI data acquisition also contribute to the uncertainties in the rough estimate of the average fiber orientation in space. Therefore, regularization methods that are needed to denoise diffusion weighted images and/or tensor data should make adequate use of neighbourhood information to utmostly preserve the underlying

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

anatomical information. Also other applications such as voxelwise statistical analysis of maps of diffusion tensor derivates and spatial normalisation algorithms would largely benefit from edge preserving interpolation tools.

This application field poses considerable statistical and numerical challenges. In particular, the massive dimension of realistic data sets does represent a computational burden if not broken down to voxelwise or sequential operations. To date, numerous concepts of denoising have been suggested at different stages of DTI data processing, ranging from anisotropic kernel methods and other adaptive filtering schemes to iterative or multilevel procedures. These methods pursue the preservation of detailed anatomical structures and the adequate consideration of the errors that follow a Rician distribution after Fourier transformation. Efforts have been spent on the issue of tensor estimation with respect to the positive definiteness and robustness as well as non-linearity of the model and mixing distributions. In view of fiber tracking, polynomial interpolation techniques and basis function approaches are in use. So far, not more than smoothing and interpolation have been melted into one processing step in order to reduce error propagation. Therefore, the main goal of this thesis is to offer a unified framework for DTI data processing with the outlook to improve fiber tracking.

To facilitate the understanding to the reader unfamiliar with the given technology, **Chapter 1** provides fundamental background information on magnetic resonance tomography and explains the biophysical principles of DTI. Apart from data acquisition the use of derived quantities is illuminated. In **Chapter 2**, the impact of noise on data quality is investigated using the non-parametric bootstrap method. The procedure allows to quantify statistical uncertainties of DTI data and derived characteristics in dependence on thermal noise and subject motion, either artificially increased or reduced by spatial smoothing and the use of a vacuum device. Also, effects of gender and age are addressed as potential confounds in group comparison studies. The extended use of bootstrap as a tool for quality evaluation of tensor derived metrics and reconstructed fiber bundles is reviewed. **Chapter 3** deals with the three-stage processing pipeline towards tractography. For the purpose of unified estimation, regularization and interpolating of the diffusion tensor field, a space-varying coefficients model (SVCM) is developed on the basis of penalized B-spline basis functions. The joint modelling of the separate regression

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

equations naturally takes into account spatial correlation by the use of basis functions which also allow the tensor estimation at any arbitrary point in space. Yet, a naive one-to-one implementation of the model formula is limited by the relatively large dimension of realistic data sets which typically contains multiple images of $(128 \times 128 \times 24)$ voxels. Tackling the problem of memory allocation, sparsity of the involved matrices of tensor product B-splines and difference penalties had to be incorporated. In addition, a sequential smoothing variant is considered as efficient alternative. A simulation study compares the novel concept to standard procedures. The chapter closes with a tutorial on the corresponding open-source software comprising both variants of the novel concept and toy data. In **Chapter 4**, wavelets are explored for their utility in DTI data processing. One- and two-dimensional examples serve for the derivation of the wavelet theory and the illustration of the multiscale property. A wavelet filter which accounts for the positive definiteness of diffusion tensors is established and examined for adaptive post-denoising of the 3d tensor field estimated on a voxel-by-voxel basis from synthetic and experimental data. Apart from replacing Gaussian kernel smoothing, wavelets could be used to substitute the B-spline basis functions in an SVCM. A proposal is made how this could be accomplished. In **Chapter 5**, the SVCM for simultaneous tensor estimation, smoothing and interpolation is finally linked to an existing tractography algorithm which is sketched in brief. Examples of reconstructed fiber bundles of the human brain are visualized and discussed. To conclude, an outlook is given on the exchange of B-spline with wavelet basis functions.

This thesis initially originates from a cooperation with the research group of nuclear magnetic resonance at the Max Planck Institute of Psychiatry, Munich. It directly picks up the work by Christoff Gössl, also developed within the project *Spatial Statistics* of the Collaborative Research Center *Statistical Analysis of Discrete Structures*. Financial support was provided by the German Science Foundation (D. F. G.).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

1 Introduction to Magnetic Resonance Imaging (MRI)

Nuclear magnetic resonance imaging deserves to be called a milestone in innovating diagnostic medicine. The chemist Paul C. Lauterbur and the physicist Sir Peter Mansfield contributed the relevant findings throughout the 1970s and were awarded the Nobel Prize in Medicine and Physiology in 2003. Medical engineering promoted the realization and continues the improvement of the magnetic resonance tomograph (Fig. 1.1) which is basically equipped with a strong magnet, a radio frequency transmitter and receiver, and gradient coils.

Figure 1.1: Patient positioned in an MR scanner for image acquisition of the human brain.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

The technique exploits the interaction of radio frequency pulses, a strong magnetic field and body tissue. It comes with considerable advantages, in particular, when compared to conventional, X-ray imaging techniques: The use of magnet resonance (MR) principles allows the generation of extremely detailed 3d images from inside the human body in a non-invasive way, without technical or health risk and without exposure to high-energy radiation. As further benefit, imaging contrasts can be adapted in a flexible manner, rendering MRI a powerful instrument to visualize and to quantify a large range of tissue properties down to a microscopic scale.

In clinical practice, MRI enables disease diagnosis and monitoring. It offers techniques that go beyond traditional imaging to show functional aspects of the investigated tissue. Diffusion tensor imaging (DTI) is one of the more advanced techniques that entered the clinical field in the 1980s. It has proven effective in studying aging processes (Moseley, 2002) and a range of neurological disorders, comprising stroke (Sotak, 2002), brain tumors, focal epilepsy, Alzheimer disease, and inflammatory diseases such as Multiple Sclerosis (Horsfield and Jones, 2002). A very promising technique for neuroscience and also clinical applications is fiber tracking, i. e. the visualization of neural fiber tracts in the brain that connect different brain regions with each other and with the myelon. This may become increasingly important in neurosurgery, when e. g. the effect of brain tumors on neighboring fiber tracts is assessed for planning of surgery (Arfanakis et al., 2006; Schonberg et al., 2006). Fiber tracking also represents a final application field of the statistical methods developed in this work (Chapter 5). In contrast, physicochemical properties such as metabolite information are generally assessed by MR spectroscopy. Moreover, functional MR imaging (fMRI), which has entered the field of neuroimaging in about 1992, has greatly advanced the understanding of normal and abnormal brain function, including sensorimotor functions but also cognitive and emotional processes. A core concept in fMRI analysis is the functional connectivity, i. e. the interaction of brain regions during a task.

At this point, DTI based tractography has become an important tool to demonstrate the real underlying anatomical connectivity. Further research aims at localizing networks of brain regions that are affected by pharmacological intervention, for example by antidepressive treatment. The outlook here is to identify, how pharmacons alter brain function

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

and to improve treatment strategies for different neuropsychiatric conditions.

This chapter is organized in two sections: first, the basics of MR physics are introduced in Section 1.1 and, second, the particular technique of DTI is explained in Section 1.2. Basic questions will be answered such as *what* is measured and *how* an image is formed. For more profound reading on both subjects, the book chapters 15 and 16 in Toga and Mazziotta (2002) can be recommended as well as the contributions in Weickert and Hagen (2006). Also, several excellent review articles provide fundamental understanding of DTI, e. g. Le Bihan (2003), Mori and Barker (1999) and Basser and Jones (2002).

1.1 Basic principles of MRI

Hydrogen is omni-present and one of the most frequent elements in the human body. The nucleus of a hydrogen atom consists of one single proton. Due to the intrinsic spin of the proton, the rotating positive charge results in a current which in turn induces a magnetic field. Hence, hydrogen protons have a magnetic dipole moment parallel to the rotating axis and can be considered as tiny magnetic needles that react to external magnetic fields. Under normal conditions, free proton spins are randomly orientated in space (Fig. 1.2, left). In the presence of an external magnetic field, the protons take on one of the two quantum mechanically possible energy levels, leading to parallel or antiparallel alignment with the magnetic field (Fig. 1.2, right). This state is called equilibrium. With increasing field strength, fewer protons are strong enough to align against the magnetic field. There is a small excess of spins aligned parallel to the external field yielding an overall net magnetization. In an MR scanning device, this magnetization vector points along the human body and is therefore referred to as longitudinal magnetization, M_z.

Yet, the protons do not comply statically with the lines of magnetic flux as suggested by Fig. 1.2, but precess, similar to tops, around the external field direction at an angular frequency. This frequency is called Larmor frequency and depends on the atom specific gyromagnetic ratio and the strength of the external magnetic field. At a magnetic field strength of 1.5 Tesla (T), commonly found in clinical MR scanners, the hydrogen spins precess at 63.9 MHz. The protons can change their spin direction if energy is added by

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

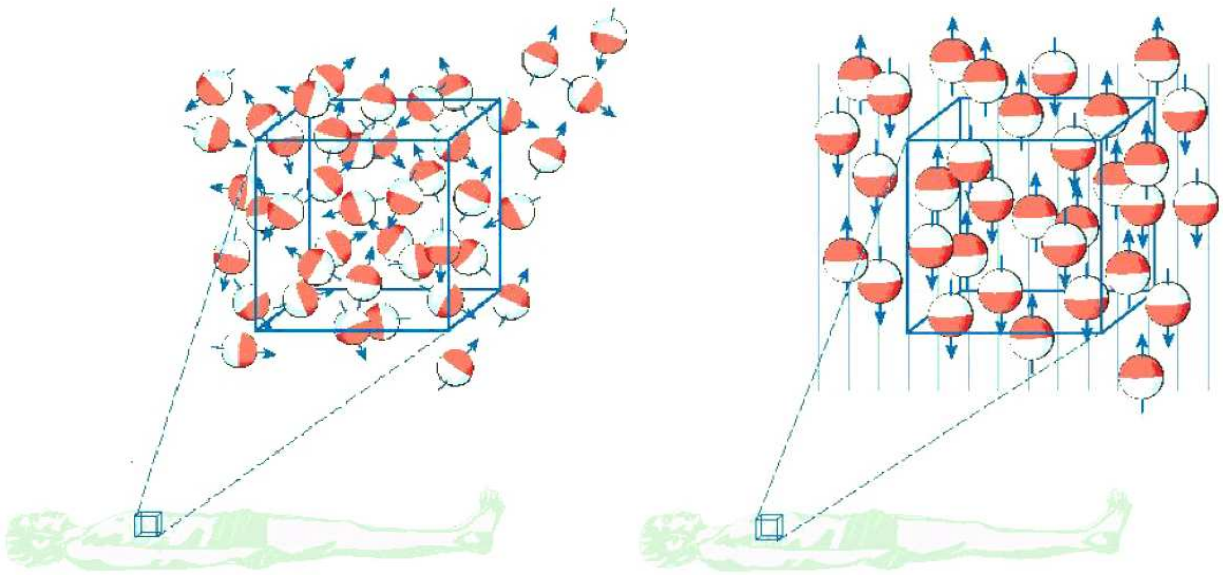


Figure 1.2: Protons are naturally in a disordered state (left), but can be aligned with an external magnetic field (right).

transmission of a suitable electromagnetic radio-frequency (RF) pulse. The RF needs to match the Larmor frequency of the magnetic spins so that the spins can couple to the magnetic component of the RF field (resonance condition) and exchange energy. The resulting spin distortion is a nuclear magnetic resonance phenomenon, giving the technique its name. After energy absorption the nuclei release this energy and return to their initial state of equilibrium. The emitted energy establishes the MR signal that is a function of contrast determining tissue parameters (proton density, relaxation times T_1 and T_2) and the machine parameters (time to echo (TE) and time to repetition (TR)), which will be explained in the following.

The most common pulse sequences employed in MR imaging are spin echo sequences. A 90° RF pulse is applied for spin excitation, i. e. deflecting the magnetic moment by 90° from the direction of the external field. As a consequence, the longitudinal magnetization is rotated into the xy-plane. Indeed, the excitation pulse causes the protons to precess in-phase, generating a transversal magnetization M_{xy}. Since the transversal magnetic moment also appears to be rotating at Larmor frequency, an MR signal voltage is induced and can be measured by a receiver station. After the shortly lasting RF excitation pulse, the transversal magnetization decreases and also the correpsonding signal intensity. The

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

free induced signal decay (FID) occurs because individual spins lose their coherence and dephase with time due to magnetic interaction among each other. Thus, the dephasing mechanism is charaterized by the spin-spin relaxation time, denoted by T_2. For the purpose of signal detection, a 180° refocusing RF pulse is applied after duration τ of the initial 90° excitation pulse, causing the spins to reverse their rotation direction. After $2\tau = TE$, the time to echo, the faster, now counterrotating spins have caught up with the slower ones and the spin echo appears in dependence on the surrounding tissue properties. For example, protons are very mobile in water and have less opportunity to exchange energy. Therefore, the amount of transverse magnetization relaxes more slowly than e. g. in fat. Water gives a stronger signal and appears brighter on the image than fat. Nevertheless, there will be some intensity loss due to local field inhomogenities.
The recovery of the longitudinal magnetization occurs with the T_1 relaxation time and typically at a much slower rate than the T_2 decay. The recovery process is driven by energy exchange of the spins with their surroundings and is therefore also named spin-lattice relaxation. Protons can efficiently interact with fatty acids (short T_1) whereas in liquids protons move too fast for notable energy transfer (long T_1). Depending on the kind of sequence used, this results in varying image brightness.

In summary, the signal intensity of the echo in a spin echo sequence is related to the square of M_{xy}, given by:

$$\mathrm{M}_{xy}(t = TE) \quad = \quad \mathrm{M}_{xy}(0)\,(1 - \exp(-TR/T_1))\,\exp(-TE/T_2), \tag{1.1}$$

where $\mathrm{M}_{xy}(0)$ denotes the transversal magnetization right after the first excitation which is equal to the longitudinal magnetization M_z at time $t = 0$ and, thus, proportional to the proton density. After later excitations, the signal intensity of the echo is essentially determined by the first two factors, where the dependency on TR describes the longitudinal magnetization rebuilt since the preceeding 90° pulse. For the moment being, it suffices to know that TR denotes the time interval before the whole pulse sequence is repeated. According to Eq. (1.1), the MR signal consists of three factors that can be down-regulated or enhanced by control of TE and TR. Different image contrasts can be obtained, named proton density weighting (short TE: 20 ms, long TR), T_1 weighting (short TE: 10–20 ms, short TR: 300–600 ms), and T_2 weighting (long TE: $>$ 60 ms, long TR: $>$ 1600 ms). Figure 1.3 gives an impression of these basic MRI modalities by means of an axial slice

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

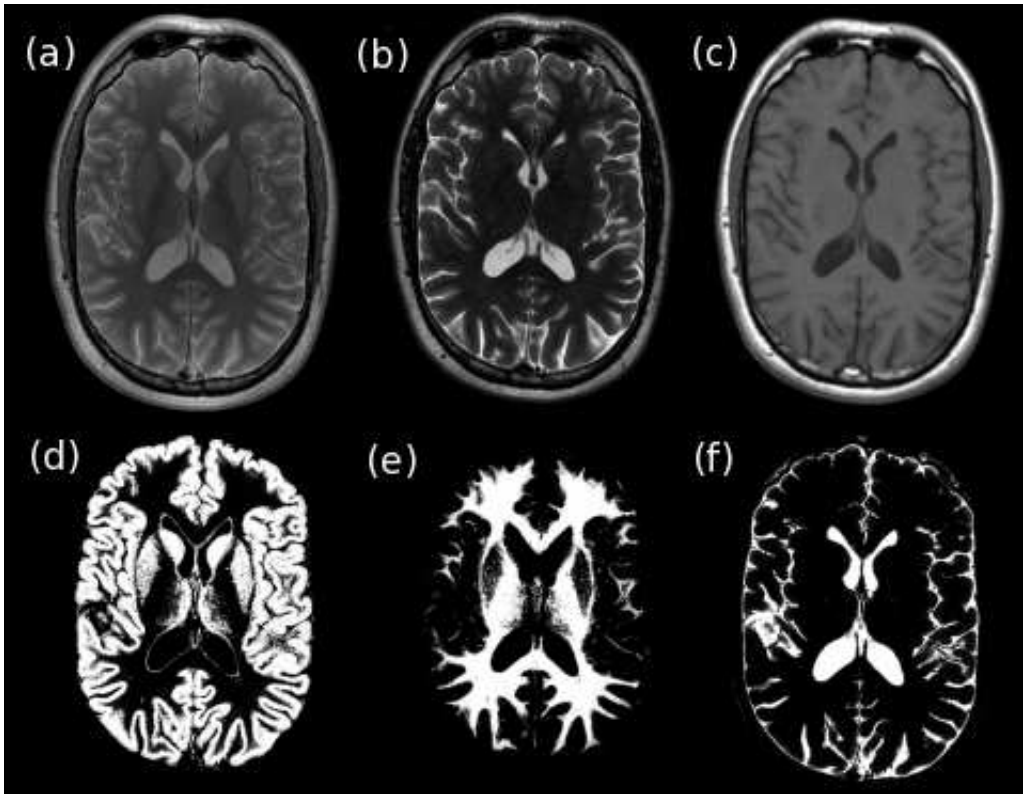


Figure 1.3: Examplary axial slice of the human brain mapped in different MRI modalities yielding (a) proton density, (b) T_1, and (c) T_2 weighted contrasts. (d)–(f) show segmented gray matter, white matter, and cerebro-spinal fluid which have different signal insensity on the upper maps (a)–(c).

of the human brain. When compared to the three compartments of cerebro-spinal fluid (CSF), gray matter (GM) and white matter (WM), the different contrasts are clearly distinguishable, e. g. CSF appears bright on the proton density and T_1, but dark on the T_2 image. The proton density contrast is determined by the number of protons per unit tissue. Since the neural fibers in white matter are wrapped with fatty myelin sheats and since fat molecules are quite space demanding, there are less hydrogen protons in WM, a smaller transversal component of magnetization, and a significantly attentuated signal compared to CSF and even GM. The proton density can be considered as a minimally T_2 weighted image.

With focus on the final MR images in Fig. 1.3, the question arises how the measured signal was spatially differentiated for image reconstruction. For the purpose of signal localization, the MRI technique makes use of additional magnetic field gradients. Slice

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

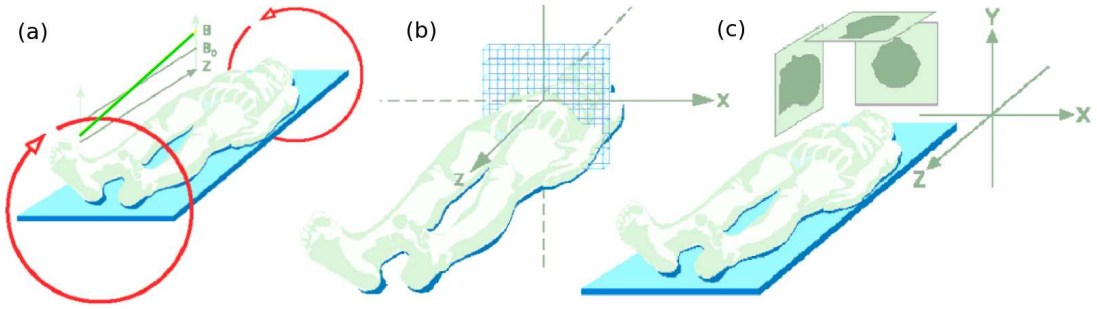


Figure 1.4: (a) Application of an additional magnetic gradient B (in green) along the z-axis serves to modulate the external field of strength B_0, thereby restricting magnetic resonance of a given RF to the protons of a specific slice; (b) a RF receiver registrates only the spin echo of protons from the blue-colored voxels; (c) further magnetic fields in vertical and horizontal direction influence the spin echo in frequency and phase, allowing to acquire a 3d volume.

selection is achieved by the application of a gradient field along the z-coordinate of the laboratory frame (Fig. 1.4 (a)). Consequently, the RF pulse tuned to the slice specific Larmor frequency only excites the respective protons (subfigure (b)). Additional phase and frequency encoding gradients permit in-plane localization of the received spin echo (subfigure (c)). Note that with decreasing size of the volume elements (voxels), fewer protons contribute to the local signal. Yet, high resolution of the final images is imperative for most applications and avoids partial volume effects, i. e. the compound of several tissue types within one voxel. To compensate for a loss in signal-to-noise ratio (SNR), excitation is repeated. For example, increasing the number of excitations from one to three results in a 73 percent increase in SNR, but also in a three-times longer scan time.
Together with the RF excitation and refocusing pulses, the sequentially applied gradients build an MRI pulse sequence. A common sequence is depicted in Fig. 1.5.

The obtained data is in the so-called k-space with radians/cm units, where k indicates the dependence on spatial frequencies k_x, k_y, and k_z defined by the gradients in x-, y-, and z-direction. Fourier transformation enables to specify data points in Cartesian coordinates. To span the full k-space line by line, a simple pulse sequence needs to be repeated after a 'time-to-repetition'-interval TR. Echo planar imaging (EPI) represents a popular variant which allows image acquisition in a rather short time. A single RF pulse suffices to generate a train of echoes by sinusoidally oscillating the spatial encoding gradients. The whole data set is then collected of the k-space in a zig-zag fashion. However, the benefits

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

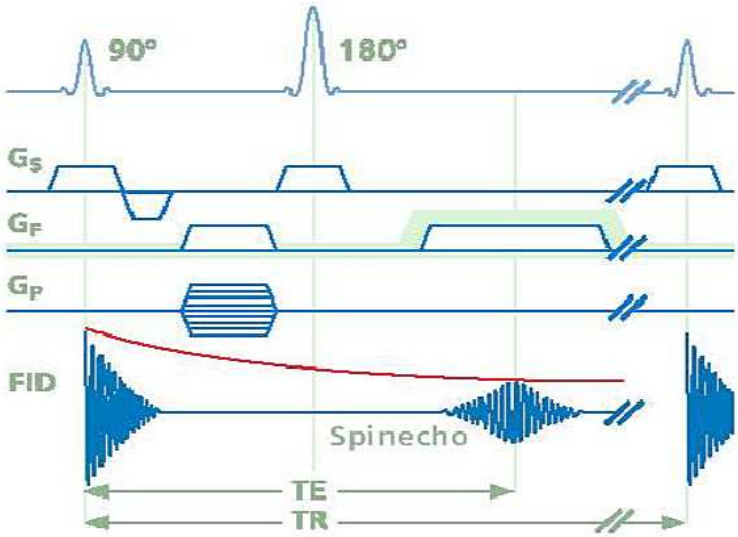


Figure 1.5: Pulse sequence timing diagram of a spin echo sequence with slice selecting gradient G_S, frequency encoding gradient G_F, phase encoding gradient G_P.

of the fast imaging time are not without additional costs. EPI is not only relatively demanding on the scanner hardware, but also extremely sensitive to image artifacts and distortions. Since EPI represents the sequence of choice in DTI, data quality is an issue to keep in mind (see Chapter 2).

1.2 Specifications in Diffusion Tensor Imaging (DTI)

Diffusion denotes the random motion of small particles that are not fixed within a sample. The process is due to their thermal energy and was discovered by the botanist Robert Brown in 1827[1]. Hence, the term Brownian motion refers also to the corresponding mathematical model which is also known as Wiener process (see Section 3.2 for more detailed notation). Primary parameters affecting diffusion encompass molecular weight, intermolecular interactions (viscosity), and temperature. In biological systems, tissue structure and compartmentalization, e. g., intra- and extracellular spaces separated by cell membranes which act as diffusional barriers, become important. In the context of brain imaging, diffusion barriers originating from neuronal and axonal membranes as well as myelin sheaths are the main determinants of the degree and directionality of diffusion

[1]Yet only in 1905, Albert Einstein corroborated the observed mobility with molecular kinetic theory.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

in brain tissue.

The molecular movement can be described best as a random walk in space, characterized by the physical diffusion constant D. Depending on the underlying sample, the magnitude of D reflects free or restricted molecular displacement. In pure liquid or samples with chaotic unorganized structures, diffusion is the same in all directions and is thus termed isotropic (compare Fig. 1.6 (a)). In contrast, diffusion can be hindered by natural barriers such as axons and myelin sheats which possibly form highly oriented and densely packed fiber bundles as present in cerebral white matter. Such a situation coerces the water molecules to preferentially move along the obstacles, yielding anisotropic diffusion as depicted in Fig. 1.6 (b) and (c). In addition to direction specific diffusivity parameters, the principal directions may be rotated against the coordinate system. Under the assumption of Gaussianity, 3d diffusion can fully be characterized by a diffusion ellipsoid, i. e. by a second order Cartesian tensor $\mathbf{D}$:

$$\mathbf{D} = \begin{pmatrix} D_1 & D_4 & D_5 \\ D_4 & D_2 & D_6 \\ D_5 & D_6 & D_3 \end{pmatrix}.$$

The tensor can be interpreted as local covariance matrix of the diffusion process and can be diagonalized by spectral decomposition. While the axes (eigenvectors) determine the principal directions, the axis lengths (eigenvalues) reflect the corresponding strength of diffusion. An isotropic diffusion ellipsoid reduces to a sphere ($\mathbf{D} = D\mathbf{I}$).

Developed in the 1980s, DTI provides local measures of water diffusion and diffusion anisotropy by measuring the mobility of water molecules along different directions. As depicted in Fig. 1.7, large gradients (gray boxes) need to be superimposed on the spin echo experiment (Fig. 1.5) to sensitize the MR images to diffusion effects in a certain direction. More precisely, a pulsed gradient is applied on either side of the 180° refocusing RF pulse according to the diffusion weighting (DW) imaging sequence in Stejskal and Tanner (1965). This affects the loss of signal intensity derived from Eq. (1.1) by an additional factor of

$$\exp(-\gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}) D),$$

where γ denotes the gyromagnetic constant of hydrogen nuclei, G the gradient pulse amplitude, δ its duration, and Δ the separation between the leading edges of the two

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

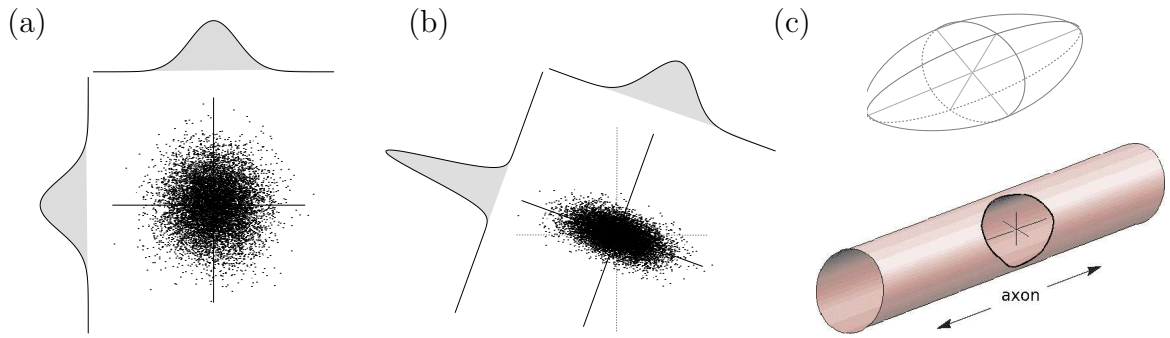


Figure 1.6: (a) Isotropic diffusion of particles due to their thermal energy shows the same dispersion in all directions as exemplified in 2d; (b) anisotropic diffusion is characterized by preferred directions; (c) the spatial organization of cerebral white matter is charactrized by highly parallel axons, determining the principal diffusion direction; thus the Brownian motion can be efficiently described by an accordingly shaped diffusion ellipsoid (tensor).

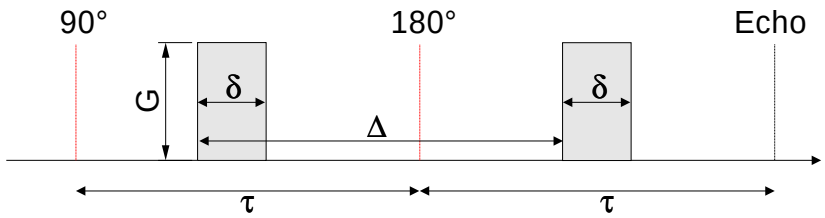


Figure 1.7: Schematic diagram of the diffusion weighting Stejskal-Tanner sequence. The purpose of the two gradients is to enhance the dephasing of the spins, thereby accelerating intravoxel diffusion processes and thus signal loss along the gradient direction.

successive gradient pulses (compare Fig. 1.7). Hence, the degree of diffusion weighted signal decay correlates with the gradient parameters, usually summarized by the b-value. The higher the value $b = b(G) = \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3})$, the stronger the diffusion weighting along the applied gradient. In areas of no diffusion, the first gradient causes the spins to dephase increasingly but the second gradient rephases them completely. With diffusion however, (rapidly) moving spins experience different fields during the two gradients so that the spins do not rephase completely. Diffusion causes a signal drop relative to the spin's velocity and the respective segment will appear darker in a DW image.

To estimate the amount of water diffusion, the relative signal decay is more important than the absolute intensity of the images which is still influenced by other contrasting mechanism, e. g. by T_2. Therefore, a reference image without diffusion weighting is required. Let S and S_0 denote the signal recorded with and without DW gradient application, respectively. Then the signal loss by a particular gradient application obeys the so-called

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Stejskal-Tanner equation (Stejskal and Tanner, 1965):

$$\frac{S}{S_0} = \exp(-bD).$$

The diffusion constant D associated with the applied gradient can then be obtained from the slope of the decay and mapped to create an apparent diffusion constant image.
For a complete picture of the spatial diffusivity distribution, diffusion weighted data needs to be acquired with many different gradient directions. In fact, six gradients in non-collinear directions build the minimum setup to determine the six unknowns in the symmetric diffusion tensors. Figure 1.8 displays the raw diffusion weighted data measured in the spatial directions as indicated by the arrows.

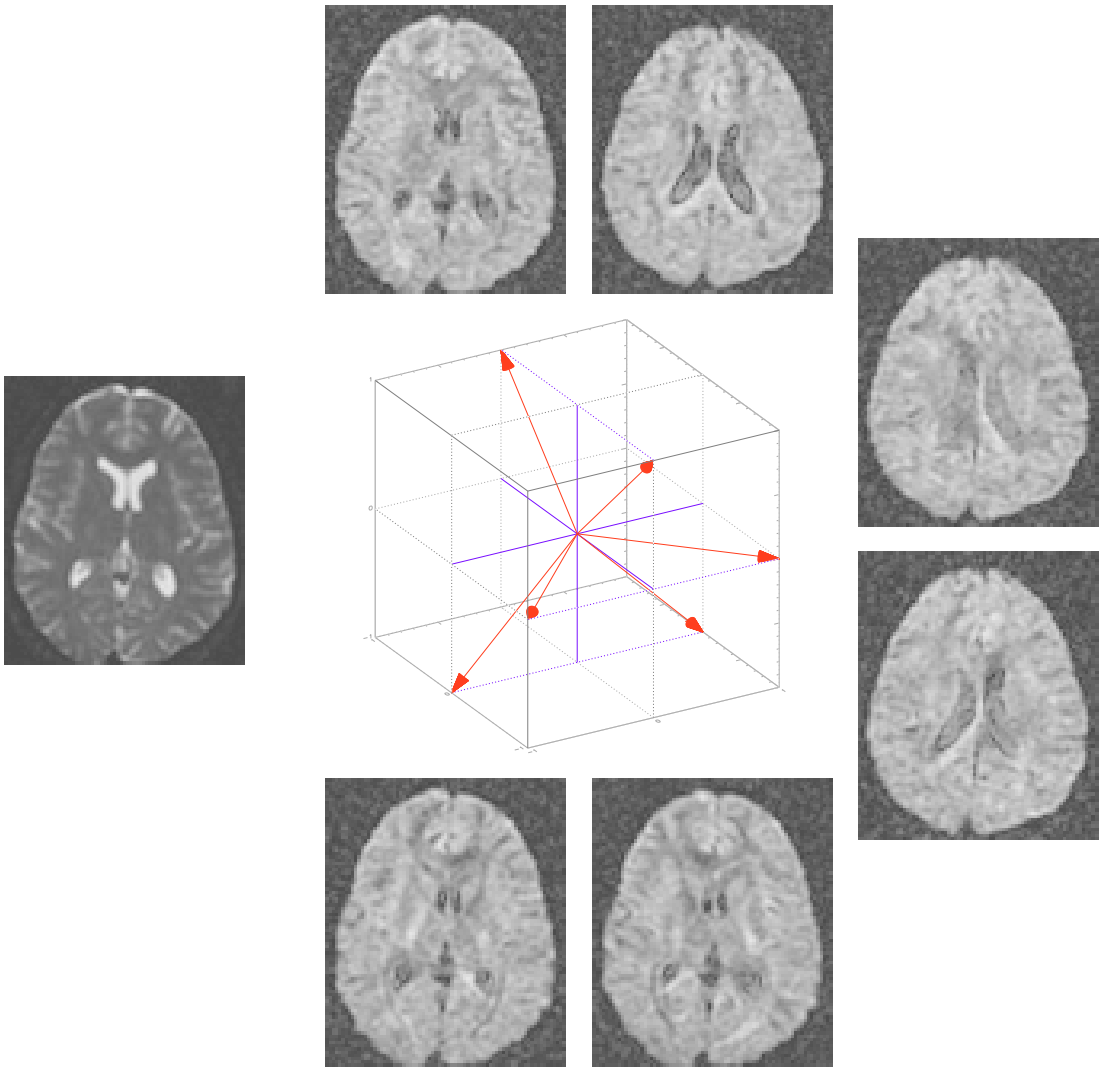


Figure 1.8: Reference ($b = 0\,s/mm^2$) and six DW images ($b = 880\,s/mm^2$) showing an axial slice of the human brain. The employed gradients spatially align with the red-colored arrows.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

The tensor elements can be estimated, for example, employing the methods developed in Chapters 3 and 4. In the standard approach, multiple linear regression is performed at each voxel independently, yielding the estimates on the left of Fig. 1.9. Physically meaningful parameters can be derived on the basis of eigenvalues and eigenvectors, resulting from tensor diagonalization. In the second column of Fig. 1.9, the upper map depicts the three possible relationships of the eigenvalues: orange voxels correspond to a cigar-shape diffusion ellipsoid with one eigenvalue clearly dominating over the others, brown voxels are associated with planar-shaped tensors with the two largest eigenvalues having very similar magnitude, and white voxels encode for almost fully isotropic diffusion where the tensor shape approaches a sphere. The lower map visualizes the principal eigenvectors, colorized in accordance to the spatial orientation: red for left-right, green for anterior-posterior, and blue for superior-inferior. The eigenvalue component serves to provide various scalar measures, e. g. the commonly used fractional anisotropy (FA, Eq. (2.1)) is based on the normalized variance of the eigenvalues. The FA map (on top of the third column in Fig. 1.9) indicates the local degree of anisotropy and, hence, how

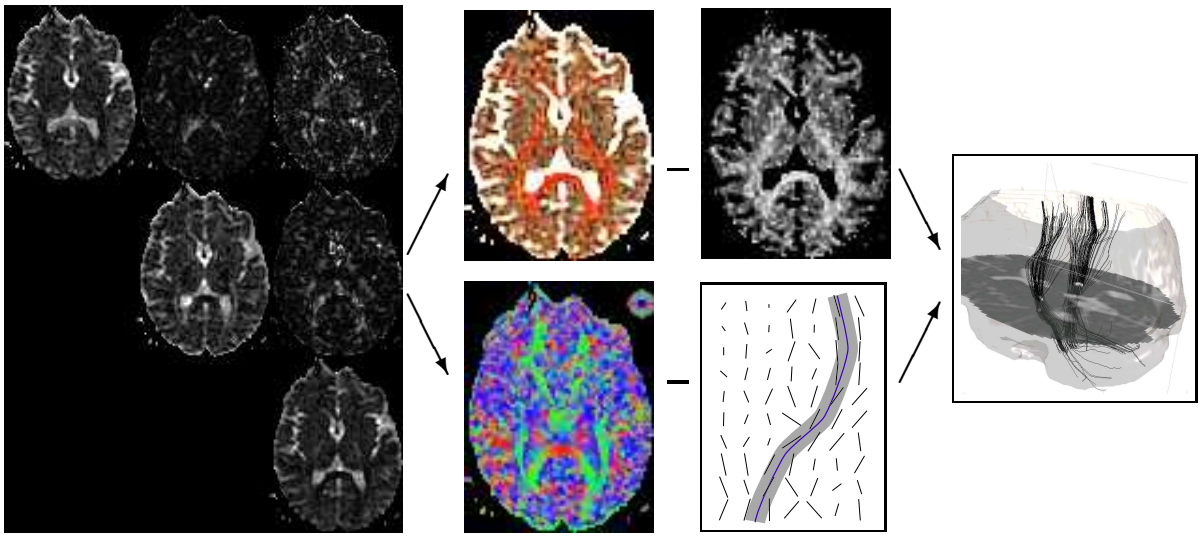


Figure 1.9: Form left to right: The tensor field can be decomposed into a map of eigenvalues and eigenvectors. While the eigenvalues (upper part) can be comprised to scalar measures of anisotropy (FA map) and thus allow to differentiate between CSF, GM and WM by appropriate thresholding, the principal eigenvectors (lower part) indicate the direction of neural fiber bundles. This allows the reconstruction of e. g. the pyamidal tract as indicated by the bundle of black fibers running from the motor cortex to the spinal cord bilaterally.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

packed and/or ordered the axonal fibers are in each voxel. Abnormal myelinization, as for example present in Multiple Sclerosis, leads to impaired diffusion directionality and to diminished values of FA in the affected regions. Also, the diffusivity might locally be less restricted, resulting in a higher mean diffusivity which is proportional to the trace of the diffusion tensor. Moreover, the degree of anisotropy can be exploited for the purpose of tissue segmentation, as it differs between white matter (> 0.2) and gray matter (< 0.15). If combined with the direction of the largest eigenvector of the diffusion ellipsoids, it forms the basis for fiber tracking as schematically depicted in Fig. 1.9. This tracking principle is based on the assumption that water tends to diffuse along fibers rather than perpendicular to them. The right most image in Fig. 1.9 gives an example of reconstructed fibers of the pyramidal tract, a major pathway of the central nervous system, transmitting motor impulses from the cortex to the spinal cord to control voluntary movement.
Note that filters for noise removal, interpolation, feature extraction are of great importance for the processing and development of DTI. This becomes particularly necessary as inherent disadvantages of DTI comprise low resolution and susceptibility to various artifacts, such as motion due to the (still) time consuming acquisition process.

The acquisition setup in Fig. 1.8 also underlies the human brain data used throughout Sections 3.5 and 4.4. All data was obtained on a 1.5 T clinical scanner (Signa Echospeed; GE Medical Systems) using a spin echo EPI sequence with TR/TE=4200 ms/120 ms. The b-value was set to $0\,\mathrm{s/mm^2}$ for the reference and to $880\,\mathrm{s/mm^2}$ for the diffusion weighted images, respectively. The field of view amounted to $24 \times 24\,\mathrm{cm}^2$. A total of seven images consisting of 24 slices with $1.875 \times 1.875\,\mathrm{mm}^2$ in-plane resolution and 3 mm thickness (1 mm gap) were obtained. The spacing helps to mitigate cross-talk effects caused by partial excitation of adjacent slices during the RF pulses.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2 Noise analysis of DTI data using bootstrap

In DTI, the raw diffusion weighted images (DWI) as acquired by an appropriate MR sequence are hardly of value for direct use. Beyond thermal and other noise that enters the data during scanning time, a chain of necessary non-linear transformations introduces further error uncertainty: At first, the magnitude is taken from the complex signal intensities and the statistics are completely formalized for this transformation only (Gudbjartsson and Patz, 1995). Yet already for the second transformation, i. e. the estimation of the diffusion tensor, different approaches exist, varying from ordinary, weighted or non-linear least squares methods to robust or variational approaches (see Chapter 3 for further discussion). In a third step, eigen decomposition of the the diffusion tensor (3×3)-matrix is performed to extract meaningful characteristics. The eigenvalues and eigenvectors build the basis for anatomically relevant measures such as the mean eigenvalue (mean diffusivity) or more complicated scalar quantities (e. g. lattice index; Pierpaoli and Basser, 1996). Apart from fruitful efforts by (Schwartzman et al., 2005) who proposed the bipolar Watson distribution for the principal eigenvector, no parametric model has been claimed so far for the tensor derived parameters. However, the uncertainties associated with estimated DTI parameters are essential components for testing hypotheses on pathophysiological conditions or for following neural fiber tracts in a probabilistic sense. For example, a group comparison study is likely to be affected if the individual data sets are of fundamentally different raw data quality and thus the corresponding FA maps (Fig. 2.1). This brings up the empirical non-parametric bootstrap methodology, a valuable tool to assess parameter uncertainties by data resampling, in particular when interest is directed towards complicated and non-linear functions of observations (Efron and Tibshirani, 1993).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

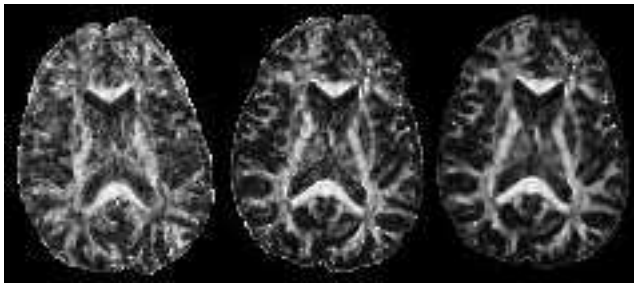


Figure 2.1: Representative slice of the FA map derived human brain data sets of different quality.

In principle, two main bootstrap variants come to mind: Either a repeated measurement design is available and resamples are generated by drawing with replacement from the observation pool or, in a more general setup that also allows for single acquisition schemes, an adequate model is fitted and artificial observations are generated by resampling the residuals. Only recently, Chung et al. (2006) pointed out that the former approach introduces substantial downward bias of bootstrap uncertainty measures and that this bias depends on the number of repetitions in each stratum, defined by the gradient directions. In constrast, the alternative approach depends on assumptions about the model to be fitted. Yet, it could be argued for the use of *model-free* bootstrapping as it may be more suitable in the presence of various sources of experimental noise: DTI data quality may be impaired by magnetic susceptibility effects, eddy currents, cardiac pulsation and head motion artifacts (Basser and Jones, 2002). With this reasoning at hand, *model-free* bootstrap has been successfully applied to characterize the parametric distributions of voxels in DTI data (Pajevic and Basser, 2003). Furthermore, deterministic fiber tracking algorithms have been endowed with local cones of fiber uncertainty (Jones and Pierpaoli, 2005) or fiber trajectory probability distribution functions to obtain probabilistic topographic maps (Lazar and Alexander, 2005). In the following, bootstrap methodology is considered for the investigation of DTI data quality, also presented in 'Assessing DTI Data Quality Using Bootstrap Analysis' (S. Heim, K. Hahn, P. G. Sämann, L. Fahrmeir, and D. P. Auer, 2004, *Magnetic Resonance in Medicine*, 52, 582–589, reprinted with kind permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc., Copyright © 2004). The book article 'Random Noise in Diffusion Tensor Imaging, its Destructive Impact and Some Corrections' by Hahn, Prigarin, Heim and Hasan (2006) can be regarded as supplementary material on this topic.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Assessing DTI data quality using bootstrap analysis

S. Heim[1*], K. Hahn[2], P. G. Sämann[1], L. Fahrmeir[3], and D. P. Auer[1]

[1]Research Group NMR, Max Planck Institute of Psychiatry, Munich, Germany

[2]Institute of Biomathematics and Biometry, National Research Center for Environment and Health (GSF), Neuherberg, Germany

[3]Department of Statistics, Ludwig-Maximilians-University, Munich, Germany

[*]*email:* sheim@mpipsykl.mpg.de

Abstract

Diffusion tensor imaging (DTI) is an established method for characterizing and quantifying ultrastructural brain tissue properties. However, DTI derived variables are affected by various sources of signal uncertainty. The goal of this study was to establish an objective quality measure for DTI based on the non-parametric bootstrap methodology. The confidence intervals (CIs) of white matter (WM) fractional anisotropy (FA) and C_{linear} were determined by bootstrap analysis and submitted to histogram analysis. The effects of artificial noising and edge-preserving smoothing, as well as enhanced and reduced motion were studied in healthy volunteers. Gender and age effects on data quality as potential confounds in group comparison studies were analyzed. Additional noising showed a detrimental effect on the mean, peak position, and height of the respective CIs at 10% of the original background noise. Inverse changes reflected data improvement induced by edge-preserving smoothing. Motion dependent impairment was also well depicted by bootstrap derived parameters. Moreover, there was a significant gender effect, with females displaying less dispersion (attributable to elevated SNR). In conclusion, the bootstrap procedure is a useful tool for assessing DTI data quality. It is sensitive to both noise and motion effects, and may help to exclude confounding effects in group comparisons.

Key words: bootstrap; DTI quality; anisotropy; reliability; motion; noise

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2.1 Introduction

Diffusion tensor imaging (DTI) is susceptible to numerous detrimental artifacts that may impair the reliability and validity of the obtained data (Basser and Jones, 2002). Thermal noise, eddy currents, susceptibility artifacts, rigid body motion, physiological pulsation flow, and hardware issues (e. g., gradient miscalibration) contribute significantly to the resulting overall noise.

While considerable research has focused on the analysis and modeling of some of these influencing factors (Anderson, 2001; Bastin et al., 1998; Skare et al., 2000b), and the development of respective correction methods (Andersson and Skare, 2002; Dietrich et al., 2001; Parker et al., 2000), relatively little effort has been spent on estimating the residual error in real data (Pajevic and Basser, 2003; Papadakis et al., 2003).

A major consequential problem is the uncontrolled propagation into derived parameters (Skare et al., 2000b; Pajevic and Basser, 2003) causing both random and systematic errors. In practical terms, this particularly affects the reliability of results from fiber tractography (Jones, 2003; Lori et al., 2002), anisotropy estimates, and, more generally, any inferences from group comparisons that may be impaired by nonidentically distributed data quality. Therefore, objective measures to assess the quality of an individual data set are urgently needed. To date, several attempts have been made to determine the uncertainty in DTI data by specific metrics (e. g., the motion artifact index (Virta et al., 1999) or a fitting error based parameter (Papadakis et al., 2003)). The signal-to-noise ratio (SNR) is the most commonly used measure of data quality; however, it is not unequivocally defined, and is highly dependent on details of the NMR diffusion sequence and encoding scheme used (Anderson, 2001; Skare et al., 2000). A further drawback is the yet unknown sensitivity of SNR to motion artifacts, which play a role in clinical trials.

Basser and Jones (2002) recently suggested that the well known non-parametric bootstrap method (Efron, 1979; Efron and Tibshirani, 1993) is advantageous for assessing DTI data quality, and estimating possible systematic artifacts. The bootstrap is a resampling technique that allows investigators to draw repeated "bootstrap samples" from real data instead of artificial data (Papadakis et al., 2003). Thus, it is possible for each individual case to approximate analytically intractable probability distributions and re-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

lated statistics (e. g., confidence intervals (CIs) and coefficients of variation (CVs)) of DT derived variables. Studies to date (Basser and Jones, 2002; Jones, 2003) have aimed to approximate analytically intractable probability distributions of DT derived variables for single voxels or regions of interest.

In this work, we extended the non-parametric bootstrap approach to generate an overall DTI data quality descriptor suitable for application studies. Toward that end, we evaluated the performance of derived quality measures by focusing on two aspects thought to directly or indirectly affect DTI data quality in a clinical setting: 1) We addressed the issue of noise systematically by artificially increasing background noise or applying voxelwise and spatial smoothing (Hahn et al., 2001). 2) We also studied the influence of both experimentally enhanced (voluntary) subject motion and reduced motion due to the use of a vacuum device. To further evaluate the relevance of data quality differences as potential confounds in group comparison studies, we tested for systematic gender and age effects.

2.2 Methods and materials

Data acquisition and processing

All DTI data sets were acquired at 1.5 T (Signa Echospeed; GE Medical Systems) using a spin-echo echo-planar sequence with TR/TE= 4200 ms/120 ms and diffusion gradients in a six noncollinear directions. A total of 20 images (three and two repetitions for $b = 880$ and $b = 0\,\mathrm{s/mm^2}$, respectively) consisting of 24 slices with $1.875 \times 1.875\,\mathrm{mm}^2$ in-plane resolution and 3 mm thickness (1 mm gap) were obtained. The effective DT and fractional anisotropy (FA) were calculated as previously described (Gössl et al., 2002), according to (Papadakis et al., 1999; Basser and Pierpaoli, 1996):

$$\mathrm{FA} = \frac{\sqrt{\frac{3}{2}\sum_{i=1}^{3}\left(\xi_i - \bar{\xi}\right)^2}}{\sqrt{\sum_{i=1}^{3}\xi_i^2}}, \tag{2.1}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

where ξ_i is ith eigenvalue of the tensor, $i = 1, 2, 3$. In addition, $\mathrm{C}_{\mathrm{linear}}$ (an anisotropy measure focusing on the tensor shape) was computed as defined in Westin et al. (2002):

$$\mathrm{C}_{\mathrm{linear}} = \frac{\xi_1 - \xi_2}{\sqrt{\sum_{i=1}^{3} \xi_i^2}},$$

where $\xi_1 \geq \xi_2 \geq \xi_3$ are eigenvalues of the tensor.

To systematically assess the suitability of an overall data quality measure, we separately analyzed brain white (WM) and gray matter (GM), thus creating a compartment with relatively high and low anisotropy. Semiautomated segmentation was performed. In a first step, we removed extracerebral tissue semiautomatically, applying a region growing algorithm and manual corrections where needed. We then separated brain WM and GM using empirically derived diffusivity and FA thresholds, as previously proposed (Cercignani et al., 2001; Chabriat et al., 1999). A cut-off of mean diffusivity $\bar{\xi} \leq 1.8 \times 10^{-3}\,\mathrm{mm}^2/\mathrm{s}$ effectively suppressed cerebrospinal fluid (CSF) (typical $\bar{\xi} = 3.0 \times 10^{-3}\,\mathrm{mm}^2/\mathrm{s}$). Voxels with FA > 0.2 and ≤ 0.2 were classified as WM and GM, respectively.

Bootstrap method

The bootstrap method (Efron, 1979; Efron and Tibshirani, 1993) is a non-parametric technique that is used to analyze variables of partly or entirely unknown distributions, such as those of FA (Skare et al., 2000b; Pajevic and Basser, 2003; Pierpaoli and Basser, 1996). The application of the bootstrap procedure yields N subsamples or resamples of a given data set by drawing with replacement, and thus provides various parameters describing the statistical properties of a measure of interest. Figure 2.2 depicts the bootstrap scheme that matches our clinical acquisition design. The s_{ij} labels the jth measurement belonging to the ith gradient direction. In order to maintain the original structure of two unweighted ($s_{0j}, j = 1, 2$) and three times six weighted ($s_{ij}, i = 1, \ldots, 6, j = 1, 2, 3$) images, the resampling is subjected to the seven subsets represented by the rows. The order within each row need not be considered, because a subsequent regression analysis is used for tensor calculation. Hence there are $\binom{n_i+k_i-1}{k_i}$ possible combinations per gradient direction $i = 1, \ldots, 6$, where n_i designates the number of repeats, and k_i is the number of drawings (here $n_i = k_i = 3$). Multiplication with the combinations from the unweighted

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

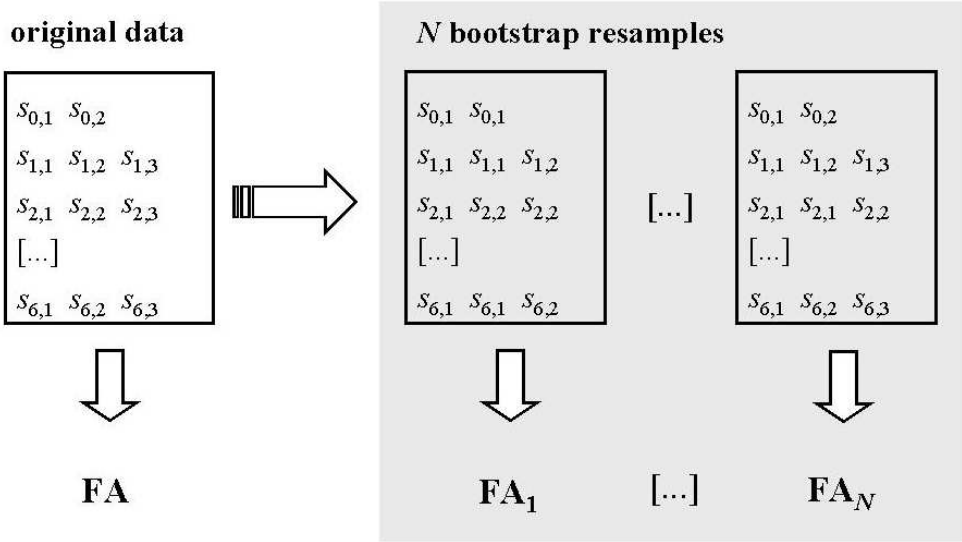


Figure 2.2: Graphic illustration of the bootstrap method adapted to the present data acquisition scheme. A (resampled) volume of 20 images is represented by one box, with rows accounting for the applied gradient direction, and columns indicating the corresponding repeated measurements. Resamples are gained by drawing with replacement of each row belonging to the original data set box. The quantity of interest (e. g., an FA map) is computed from each bootstrap resample. These replicates can be transferred into one single map of voxelwise CIs.

case ($n_0 = k_0 = 2$) yields a total of $3 \cdot 10^6$ possible new data sets. Evidently, the reliability of the bootstrap method depends on the number N of resamples and the number n_i of equivalent observations.

For the purpose of inferring the overall uncertainty contained in a single data set, $N = 100$ resamples are generated, yielding a voxelwise distribution of the respective variable (such as FA, C_{linear}, or mean diffusivity) from which the width of the 95% CI is extracted as the key feature. This first information of voxelwise reliability is then comprised in a histogram whose mean, peak position, and height provide variables of the individual data set quality. Normalization to the corresponding number of studied pixels allows a direct comparison of each histogram's shape.

To achieve comparability of the observed data quality between different acquisition or processing schemes, and between various DTI derived metrics, CVs were voxelwise extracted from the bootstrap derived distributions.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Noising

The mean SNR of a certain b-value image is defined by (Bastin et al., 1998; Parker et al., 2000):

$$\mathrm{SNR}_{b,i} = \frac{\overline{S_{\mathrm{ROI}}}}{\sigma n_i^{-0.5}},$$

where the mean magnitude signal S within a region of interest (ROI), calculated over the $n_i, i = 0, \ldots, 6$, available repetitions for a given b-value, determines the nominator.
The denominator describes the Gaussian noise of the signal, which is estimated in accordance with Gudbjartsson and Patz (1995):

$$\sigma = \frac{\sigma_{\mathrm{background}}}{0.66}. \tag{2.2}$$

The background ROI covered four bars that were positioned far enough away from both the skull and the image margins to avoid contamination due to artifacts. Modification with the factor 0.66 accounts for the Rayleigh distribution, while multiplication with $n_i^{-0.5}$ adjusts to repeated measurements. The SNR of one volume is then determined as the averaged $\mathrm{SNR}_{b,i}$, $i = 0, \ldots, 6$ referring to our data acquisition scheme.

For the purpose of investigating the effect of thermal noise, 15 control data sets from an unrelated study (seven males and eight females, mean age = 28.47 years, range = $22 - 35$ years) recorded under normal conditions were additionally impaired. Artificial noise was simulated from $N(0, \alpha\sigma_0^2)$, with σ_0 being an estimate of the present underlying Gaussian noise (cf., Eq. (2.2)). Following Parker et al. (2000), the generated noise is added to the magnitude signals, such that the resulting negative intensities are set to zero. This operation leads to small deviations from the intended percentage increase of $\alpha \cdot 100\%$ for the involved SNR range in the brain. With the aim of simulating pseudo-realistic noise levels, the maximal increment was chosen to reflect the spread of usual noise derived from a control group ($N = 24$), i. e., mean regular noise plus twice the standard deviation (SD). Table 2.1 details the resulting noise degrees and averaged SNRs for different regions of interest. The original mean S_{brain} averaged to 123.41 ± 12.99 SD, and the original background noise reached 10.30 ± 0.69.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Effective percentual increase of noise	0%	10%	20%	30%	40%
Brain	19.80 ± 2.59	18.01 ± 2.35	16.73 ± 2.18	15.35 ± 2.01	14.34 ± 1.88
White matter	18.28 ± 2.20	16.62 ± 2.00	15.45 ± 1.86	14.17 ± 1.71	13.24 ± 1.60
Gray Matter	22.57 ± 2.94	20.52 ± 2.67	19.07 ± 2.47	17.49 ± 2.28	16.35 ± 2.12

Table 2.1: Mean SNR ± SD of three compartments corresponding to different levels of percentual noise increment.

Smoothing

Before the tensor calculation was performed, noise reduction was achieved by means of a non-linear spatial filter operating on the magnitude signal field. The filter consists of a chain of non-linear 3d Gaussian filters (Winkler et al., 1999) with chain parameters adapted to the specificities of DTI data (Hahn et al., 2001). The smoothing procedure is particularly appropriate for accounting for both the inherent discontinuities of DTI, such as edges in anisotropy values or in their main directions, and the appreciably curved smooth regions between such edges. An optimal approximation for this data type was obtained by three iterations of the chain, where both the linear spatial and the non-linear signal dependent windows of the filter were parametrized as proposed previously (Hahn et al., 2001). This filter construction estimates the local mean of the magnitude images, and is statistically robust to the minor skewness and heteroscedasticity of the corresponding Rician noise distributions at locally low SNR (Gudbjartsson and Patz, 1995).

Motion experiments

For this small experimental study, six volunteers (three females and three males, mean age = 26.8 years, range = $22 - 31$ years) were investigated under three different conditions during one session. To reduce motion as much as possible, we performed the DTI acquisition while the subjects were held fixed by a vacuum cushion. Then, each subject was instructed to lie as still as possible in the scanner with standard fixation but no vacuum cushion (i. e., the standard clinical setting). During the last DTI acquisition, the subjects were asked to yawn, chew, shake, or nod their head in order to voluntarily induce

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

motion in the x-, y-, and z-directions. The subjects gave their consent in accordance with the Declaration of Helsinki, and the scanning protocol was approved by the institutional ethics committee.

Gender and age effects

Confounding effects from data quality differences may have important practical implications when they are uncontrolled in group comparison studies, as is currently done in neurobiologic and clinical studies. To test whether the reported inconsistencies regarding gender differences in FA (Virta et al., 1999; Sullivan et al., 2001; Westerhausen et al., 2003) and the more consistently observed age effects (Abe et al., 2002; O'Sullivan et al., 2001) may be related to data quality, we applied a bootstrap analysis to a larger data sample from 45 controls (21 males and 24 females, age $= 48.5 \pm 18.6$ years, range $= 22 - 84$).

Statistical analysis

We examined the results of the experiments by multivariate repeated measurements analysis, using the mean CI, peak position, and height as response variables (unless otherwise indicated). The results from multivariate analysis of variance (MANOVA) were interpreted by means of Wilks' Lambda, followed by univariate F-tests. The significance level was set to 0.05. Pearson's coefficient of correlation r with uncorrected P-values for two-tailed tests served as the interrelation measure. In case of small sample sizes however, a Spearman rank analysis was performed. The reported values are given as the mean and SD.

2.3 Results

Noise

As expected, the increasing data corruption due to increasing noise levels led to reduced certainty of the calculated anisotropy measures FA and C_{linear}), as reflected by a right

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

shift and flattening of the normalized histogram over the respective CIs (Fig. 2.3 for FA). The results for C_{linear} were similar to those for FA, but less distinct. Therefore, for the remaining analysis we focused solely on FA. The difference between the five noise levels was significant in both WM (Wilks' Lambda = 0.003, $P = 0.001$) and GM (Wilks' Lambda = 0.002, $P = 0.001$). All dependent variables (e. g., the mean, and peak position and height) contributed significantly, and even the lowest noise level revealed a significant difference from the original. Inspection of the original data sets from 44 probands revealed a strong interrelation between the three bootstrap parameters and the SNR measured in WM voxels ($r = -0.71$ mean CI, -0.80 peak position, and 0.69 peak height), which explained about $35-50\%$ of the variance of each histogram derived metric. These correlations did not alter much when they were taken from GM voxels ($r = -0.59$ mean CI, -0.77 peak position, and 0.50 peak height).

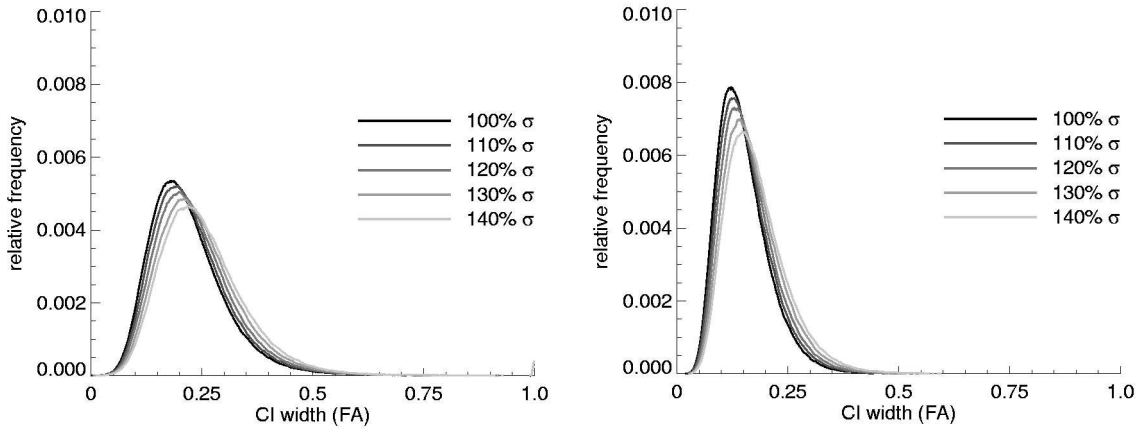


Figure 2.3: Averaged normalized histograms ($N = 15$ healthy volunteers) of the CIs for FA of WM (left) and GM (right) voxels for various degrees of additional noising, given as a percentage of the original background noise level σ. Note that the smaller CIs in GM do not reflect higher precision (compared to WM) without normalization to the lower FA. The respective CVs show the expected lower precision in GM (see text).

Smoothing

In general, edge-preserving smoothing improved original data quality by left shift and narrowing of the respective histograms (Fig. 2.4). Statistical analysis verified the visual finding of a distinct improvement from edge-preserving smoothing (WM: Wilks' Lambda

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

$= 0.004$, overall $P < 0.0001$; GM: Wilks' Lambda $= 0.006$, overall $P < 0.0001$). Simple contrasts of the smoothing or noising level to normal gave similarly significant P-values for each response variable. It is noteworthy that smoothing achieved uniformly improved data quality with less dispersion than the standard data sets, largely irrespectively of the applied noise levels (Fig. 2.4). Moreover, smoothing led to systematically lower FA values with mean FA of WM voxels, assuming 0.41 ± 0.013 in the unsmoothed case, and 0.37 ± 0.012 in the smoothed case (paired t-test: $P < 0.0001$, $N = 45$). As expected, the correction of the noise dependent bias was slightly larger for the GM compartment (FA 0.13 ± 0.005 unsmoothed vs. 0.11 ± 0.004 smoothed).

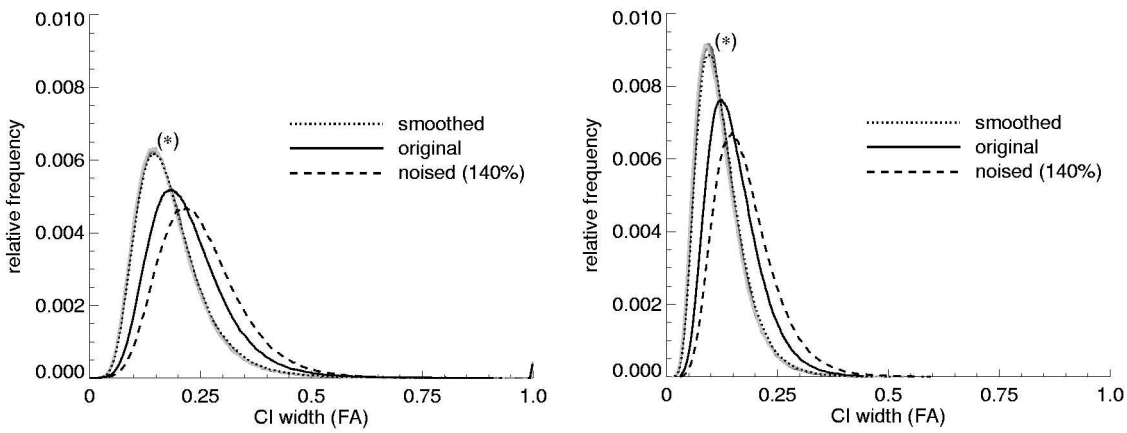


Figure 2.4: Averaged histograms ($N = 15$) of the CIs for FA of WM (left) and GM (right) representing the effect of smoothing and noising compared to the unchanged original data. When the artificially noised data was smoothed, the extra dispersion from added noise levels (up to 40%) was completely reversed. This postprocessing step lifted the corrupted data to the same quality level as achieved when the original data sets were smoothed. Corresponding histograms (gray lines) are superimposed in (*).

Motion

The graphic display in Fig. 2.5 shows an evident effect of additional movements on data quality, with greater characteristics in WM than in GM, thus making any further quantification dispensable in this regard. In contrast, the potential influence of the vacuum cushion is of interest, particularly with respect to necessary changes in the clinical protocol. Even though this study involved a small sample and well instructed, young, healthy

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

volunteers, the vacuum cushion still improved data quality somewhat (Fig. 2.5). To account for the sample size, a non-parametric Wilcoxon test was used, and the results confirmed a significant improvement in GM ($P = 0.031$ peak position and height). There was a trend for improvement in WM ($P = 0.063$ peak height, $P = 0.094$ mean CI). Concerning data sets containing strong motion artifacts (dotted line in Fig. 2.5), the question arises as to whether non-linear smoothing can rescue low-quality data. However, even sophisticated edge-preserving smoothing is not a remedy for such motion artifacts, and it improved the mean 95% CI only marginally ($P = 0.031$).

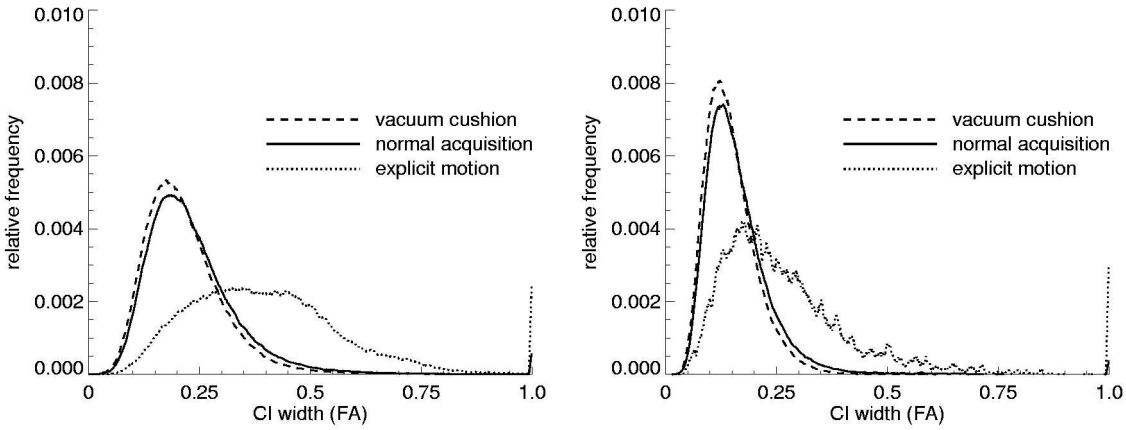


Figure 2.5: Averaged histograms of the CIs of FA of WM (left) and GM (right) voxels, obtained from six individuals undergoing DTI acquisition under three conditions: normal head padding, vacuum cushion, and no restraint and voluntary motion.

Anisotropy

Bootstrap derived CIs were similarly affected by noise and motion effects in both compartments, which differ largely in the respective mean FAs. We tested for interrelations between data quality (mean CI) and both FA values and SNR within each compartment before and after smoothing. In the original data, the FA and mean CI were positively but weakly correlated ($r = 0.318$, $P = 0.035$ in WM, $r = 0.338$, $P = 0.025$ in GM, $N = 44$ after an outlier was discarded). SNR was strongly and negatively correlated with mean FA in WM and GM ($r = -0.606$ and $r = -0.663$, $P < 0.0001$, respectively). The mean FAs from the smoothed data were no longer interrelated with mean CI or SNR.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Relative errors of DTI metrics

In a sample of 15 healthy volunteers, we additionally calculated the mean coefficient of variation (CV), and the results revealed a higher CV for FA in GM compared to WM (GM: 0.25 ± 0.01, WM: 0.15 ± 0.01). In both compartments, FA was less precise than mean diffusivity (GM: 0.10 ± 0.10, WM: 0.09 ± 0.04), but more precise than C_{linear} (WM: 0.30 ± 0.02, GM: 0.44 ± 0.01). All mutually paired t-tests gave $P < 0.0001$. As expected, the CV provided data quality similar to that obtained with histogram derived measures, with the following coefficients of correlation for FA: WM ($r = 0.907$ mean CI, 0.926 peak position, -0.852 peak height), and GM (0.944 mean CI, 0.844 peak position, -0.952 peak height, $N = 15$).

Gender and age effects

A MANOVA (one-factorial: gender; mean CI, peak position and height from GM and WM as dependent variables; age and total number of brain pixels as covariates; $N = 45$) revealed a distinct gender effect (df = 6, Wilks' Lambda = 0.523, $P < 0.0001$). Age, but not brain size, exerted a covariate effect (Wilks' Lambda = 0.600, $P = 0.004$). Univariate F-tests showed that all data quality measures contributed significantly, with mean CI, peak position and height in WM having larger effects ($\mathrm{F} = 15.63, 20.99, 19.15$) compared to GM ($\mathrm{F} = 6.69 - 8.22$) (Fig. 2.6).

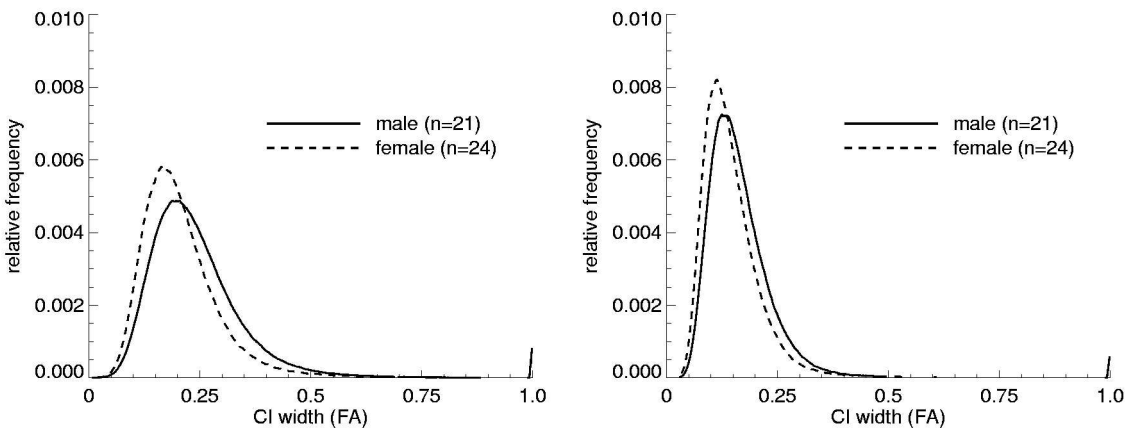


Figure 2.6: Averaged histograms of the CIs of FA of WM (left) and GM (right) voxels, illustrating the effect of gender for a sample of 45 probands.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

To further assess the potential causes of the observed gender effect, we examined the SNR differences. Female subjects displayed larger SNRs than males in both WM (20.21 ± 1.55 vs. 17.20 ± 1.52, two-tailed t-test, $P < 0.0001$) and GM (24.35 ± 1.70 vs. 20.71 ± 1.49, $P < 0.0001$), whereas background noise was similar between groups. Consequently, the model was extended for SNR as a further covariate, which eliminated the gender difference (Wilks' Lambda $= 0.897$, $P = 0.675$), whereas age and SNR covaried significantly.

2.4 Discussion

This study demonstrates the usefulness of the bootstrap method for deriving an automated and global quality measure for DT based variables. We found that histogram derived metrics of the CIs of anisotropy measures were sensitive to minor changes in noise levels or motion. Bootstrap derived CVs further allowed us to compare the precision between DT variables and tissue compartments. Moreover, significant age and gender effects on data quality were observed, highlighting the need to control for related confounds in group comparison studies.

Although the bootstrap approach is highly sensitive to several factors that frequently impair data quality, the underlying cause of the changes in histogram shape and related measures cannot be determined. The bootstrap resamples and derived histogram metrics are sensitive to test-retest instability irrespective of its source (e. g., noise, between scan motion, or technical scanner instability).

To explore the respective contributions of some defined types of noise and motion on this quality measure, we varied the noise and motion. This revealed the expected major detrimental effect of noise on data quality in a linear fashion over the studied noise levels, which was confirmed by a direct correlation. However, denoising by means of an edge-preserving smoother appears to behave non-linearly. The CI of the obtained FA values from the five differently noised data sets were reduced to an identical level irrespective of the initial noise level, with Rayleigh background noise being decreased to approximately 20% of the intrinsic noise level with the use of the described spatial filter chain. The efficacy of the denoising was further shown by the cancellation of the previously strong interrelation between the FA and both data quality and SNR. Moreover, smoothing

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

reduced the alleged noise induced bias of FA values (Skare et al., 2000b; Pierpaoli and Basser, 1996) by lowering the respective values, especially in less reliable GM.

By using simple motion-varying experiments, we were able to show that bootstrap metrics are indeed sensitive to motion dependent quality impairment. Admittedly, voluntarily induced motion mirrors clinically relevant motion only roughly, in both character and extent (which was exaggerated by some volunteers). While this does not allow for detailed analysis that would necessitate dedicated simulation experiments, the observed partial improvement in data quality measures from the use of a vacuum cushion is remarkable and shows the sensitivity of the data quality assessment. In highly motivated young volunteers, the use of a vacuum cushion (as opposed to standard padding) cannot be expected to substantially reduce head motion. Nevertheless, in this small study sample ($N = 6$), a noticeable improvement was identified for peak position shift and dispersion in the GM compartment.

In comparison with previously described techniques to assess data quality, the chosen bootstrap method has several advantages. One distinctive feature of the bootstrap approach is that one can directly assess the error size being propagated into any DT derived variable, and thus achieve comparability between variables and tissue compartments, and (potentially) scanners and acquisition schemes. The relative errors expressed as CV clearly showed that FA was twice as unreliable as mean diffusivity, but about twice as precise as C_{linear}. Such information is clearly helpful for selecting appropriate variables in application studies, and is not obtained from SNR or χ^2.

Our adaptation of the bootstrap procedure shares the noise sensitivity with a few alternative indices of data quality assessment. Papadakis et al. (2003) published a modified version of the known fitting error χ^2 in order to filter voxels with unreliable signal information. The presented bootstrap method can also be used as a filtering tool, and it has the advantage of being empirically validated instead of relying on simulated data. Moreover, while χ^2 is sensitive to both the goodness-of-fit and the SNR, the impact of motion remains unclear.

The proposed motion artifact index allows DTI corruption to be determined conditional on uncorrected nonrigid motion (Virta et al., 1999). Though no explicit investigation was spent on the sensitivity of the motion artifact index to SNR and other sources of

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

instability, its primary goal was not to provide an integral quality measure as sought-after in clinical neuroscience applications. All four presented bootstrap derived quality measures, namely the three histogram-metrics mean, peak position and height, and the average coefficient of variation, seem useful in this respect as they successfully detected both noise impairment and motion effects. Beyond the intended applications in clinical and neurobiological studies, these data quality measures are well suited for comparing the performance of various DTI acquisition schemes and scanners, and the effects of field strengths and complex correction methods (Andersson and Skare, 2002).

As regards methodological issues and limitations, some critical points should be considered. As is obvious from the Monte Carlo simulations, the bootstrap results are dependent on several factors: 1) the seed of the random number generator, 2) the size of the sampling pool, and 3) the number of generated subsamples. Whereas factor 3 basically increases the offline computational effort, the size of the total sample pool (factor 2) can only be increased by the acquisition of more repeat scans, which inevitably extends the total scanning time. Nevertheless, reliability investigations demonstrate that bootstrap derived indices of data quality are highly precise, as indicated by the CV between 0.011 and 0.039 (20 bootstrap runs with 100 resamples).
The observed age and gender effects are noteworthy because they highlight potential confounds in simple group comparison studies, which may explain certain controversial findings (Virta et al., 1999; Sullivan et al., 2001; Westerhausen et al., 2003). Higher SNR and overall better data quality in female brains compared with male brains confirm and extend earlier results from a smaller sample (Virta et al., 1999). While our study did not attempt to separate motion from noise sources leading to poorer data quality in male DTI data sets, the results underscore the need to account for these factors (Virta et al., 1999) if biologically meaningful gender differences are to be ascertained. Similarly, one would have to separate an age related decline (Virta et al., 1999; Abe et al., 2002; O'Sullivan et al., 2001) in FA from the indirect effects of age related data quality reduction.

It is conceivable that future modifications and extensions of the bootstrap technique will increase the application range. First, as an alternative to the non-parametric regression bootstrap, the so-called residual sampling or parametric regression bootstrap (Freedman, 1981) appears to be a useful and complementary tool. It allows bootstrap-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

ping to be performed with a fixed design of encoding directions, and does not require repeated measurements, as commonly used for high-angular-resolution diffusion weighted imaging (HARDI; Frank, 2002). On the other hand, more stringent assumptions on the error variables of the regression model must be made, and it must be determined whether these assumptions are consistent with DTI experiments. Second, extensions of the bootstrap technique to recently developed fiber tracking algorithms (Gössl et al., 2002; Basser et al., 2000) would be of great interest (Jones, 2003). Currently, line propagation-tracking algorithms work deterministically by providing "point" estimators for fiber curves without evaluating the degree of uncertainty of tracking results. Voxelwise assessments of the uncertainty of eigenvalues and eigenvector directions have been achieved by bootstrap methods (Jones, 2003), but it remains unclear whether such information from neighboring voxels can be successfully combined within existing tracking algorithms. The Bayesian state space model (Gössl et al., 2002) may offer an alternative approach for deriving probabilistic credible regions for fiber tracts. Fully Bayesian tracking techniques, which allow samples of tracks to be drawn based on Markov chain Monte Carlo (MCMC) methodology, may be feasible, although they are computationally much more demanding. As discussed in Hastie et al. (2001, p. 235), there is a close relationship between the bootstrap and Bayesian inference, indicating that one could think of the bootstrap distribution as a "poor man's" Bayes posterior.

Acknowlegdements

The authors thank Rosa Hemauer, Reinhold Borschke, and Elke Schreiter for performing the data acquisition, and Uwe Wolff for preparing the whole brain masks. We gratefully acknowledge financial support from the Collaborative Research Centre 386 *Statistical Analysis of Discrete Structures*, sponsored by the German Science Foundation (D. F. G.).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3 Space-Varying Coefficients Model using 3d B-splines

In this chapter, the current analysis framework of separate voxelwise regressions is reformulated as a 3d space-varying coefficient model (SVCM) for the entire set of diffusion tensor images recorded on a 3d voxel grid. Beyond an improved estimation of the diffusion tensor field, the approach casts the three-step cascade of standard data processing (voxelwise regression, smoothing, interpolation) into a unified framework using penalized B-splines. Thereby strength is borrowed from spatially correlated voxels to gain a regularization effect right at the estimation stage. Two SVCM variants are conceptualized: The first one is a full tensor product approach (Section 3.3.2), that shares the theoretical model with array regression (Eilers et al., 2006) but takes advantage of the sparsity of the spatial arrays involved. The second one (Section 3.3.3) basically adapts the 'new smoothing spline' in Dierckx (1982), thus reducing the 3d (or higher-dimensional) problem to a sequence of one-dimensional smoothers. The resulting sequential approximation renders the SVCM numerically and computationally feasible even for the huge dimension of the joint model in a realistic setup.

The main part of this chapter consists of a preliminary and extended version of an article accepted for publication in *Computational Statistics and Data Analysis* (S. Heim, L. Fahrmeir, P. H. C. Eilers and B. D. Marx, 2007, '3d Space-Varying Coefficient Models with Application to Diffusion Tensor Imaging', *Computational Statistics and Data Analysis*, reprinted with kind permission of Elsevier). After a brief motivation in Section 3.1, essential background information on DTI is given along with an introduction to current postprocessing methods (Section 3.2). Section 3.3 deals with the conceptual framework

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

composed of the above two estimation techniques. In deviation from the paper, Section 3.3.1 adds substantial information on multidimensional B-splines, and Section 3.3.3 is elaborated in full detail, encompassing a precise comparison to the tensor product variant (Section 3.3.2). The performance of the proposed methods for DTI analysis is explored through a simulation study in Section 3.4. The Supplement 3.4.1 comprises further results from simulations beyond the minimum acquisition design, presented at the 12th Annual Meeting of the Organization for Human Brain Mapping (Florence, Italy; 2006). In Section 3.5 the application to real DTI data is examined. Discussion and conclusions are presented in Section 3.6. Section 3.7 provides some additional explanatory remarks on remaining drawbacks. Technical details of the implementation have been deferred to Section 3.8 with a comprehensive description of the associated software package in Section 3.9, also not included into the article.

3.1 Introduction

The presented work results from collaborative research in neuroscience with focus on diffusion tensor imaging (DTI), one of the advanced magnetic resonance techniques (Basser and Jones, 2002; Mori and Barker, 1999). In DTI, the term diffusion refers to the translational motion of water molecules that is also known as Brownian motion. This physical process results from the thermal energy carried by these molecules. In 3d tissue samples, diffusion can be described by a positive-definite (3×3)-matrix $\mathbf{D}$, called the diffusion tensor. As $\mathbf{D}$ is symmetrical this matrix consists of six unknown elements which we are interested in estimating. DTI allows the anisotropy of water diffusion to be studied, which in turn allows to recover anatomical connections between brain regions. This ultimately intended high-end application of DTI is also referred to as fiber tracking (Mori and van Zijl, 2002; Le Bihan, 2003).

This article pursues two primary goals motivated by the manifold challenging issues in DTI. The first goal is to extend current 1d and 2d methodology for space-varying coefficient models (SVCM) to the 3d case. Concerning the implementation, the use of the recently published array regression (Eilers et al., 2006) comes into question but poses substantial memory problems. We circumvent this storage issue and the computational

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

burden by exploiting the sparsity of the involved matrices. The developed method can equally be applied to many other research fields engaged with discrete multidimensional mesh data. The second goal is to analyze simulated and real DTI raw data using the 3d SVCM which represents an appealing unified framework. The novelty of the approach lies in its ability to simultaneously estimate, regularize and interpolate the diffusion tensor field, thereby reducing error accumulation. For the first time, information from neighbouring spatial elements is considered right at the stage of estimation, clearly distinguishing our approach from reported methods.

DTI suffers from two major limitations: Raw data is inherently artifact-prone and recorded at low resolution on a discrete 3d grid of voxels, indexed by $s = 1, \ldots, n$. In a realistic setup, a typical image comprises $n_1 \times n_2 \times n_3 = 128 \times 128 \times 24$ voxels, resulting in a total of $n = 393,216$ voxels. A complete data set consists of multiple images recorded under different conditions and can be transformed to repeated continuous measurements $y_i(s), i = 1, \ldots, r$, at each voxel s. These measurements can be related to the vector $\boldsymbol{\beta}(s) = (\beta_1(s), \ldots, \beta_6(s))'$ of unknown elements in $\mathbf{D}(s)$ through a regression model

$$y_i(s) \;=\; \mathbf{x}_i{}'\boldsymbol{\beta}(s) + \varepsilon_i(s)\,, \quad \varepsilon_i(s) \overset{\text{iid}}{\sim} N(0, \sigma^2), \quad i = 1, \ldots, r. \tag{3.1}$$

The covariate vector $\mathbf{x}_i$ is determined by the design of the DTI experiment.

For the purpose of assessing the diffusion tensor $\mathbf{D}(s)$, several "model-free" approaches such as high angular resolution imaging (HARDI; Frank, 2002), Q-ball imaging (Tuch, 2004), and persistent angular structure MR imaging (Alexander, 2004) are published. Yet, the majority of efforts has been invested in rendering the basic diffusion tensor model (3.1) biologically more realistic on the one hand and in improving its estimation on the other hand. Current methods for estimating the coefficients $\boldsymbol{\beta}(s)$ and, thus, the quantity of interest $\mathbf{D}(s)$ apply a regression technique in each image voxel separately, reaching from standard least squares (Basser et al., 1994) to more sophisticated techniques: While some authors consider diffusivity contributions from different tissue compartments (cerebrospinal fluid, gray and white matter) by a weighted tensor model (Basser and Jones, 2002) or conceptualize contributions from two non-exchanging pools of water molecules by a dual-tensor model (Clark and Le Bihan, 2000), others incorporate extra model terms accounting for intra-cellular diffusivity in white matter leading to non-Gaussian regression

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

(CHARMED; Assaf et al., 2004). In addition, attempts towards robust tensor estimation have been made, e. g. by replacing the commonly used (non-)linear least-squares tensor fitting with the Geman-McLure M-estimator (Mangin et al., 2002) or with iteratively reweighting least-squares regression to identify potential outliers and subsequently exclude them (RESTORE; Chang et al., 2005). Considerable effort has been invested to preserve the positive definiteness constraint of the diffusion tensor (Tschumperlé and Deriche, 2003; Wang et al., 2004; Koay et al., 2006).
However, all existing methods are two-stage since correlation between adjacent voxels is taken into account in a postprocessing step, if at all. Though two-stage procedures are convenient with respect to computation time and memory storage, error propagation is often uncontrolled. The various concepts for post-hoc regularization of the tensor field comprise Gaussian (Gössl et al., 2002) and anisotropic kernels (Ding et al., 2005; Mishra et al., 2006) as well as other adaptive filtering schemes (Samsonov and Johnson, 2004). To our knowlegde, this is the first approach which directly incorporates spatial information into the estimation process.

We propose to connect the seemingly unrelated regression models in Eq. (3.1) to a joint model for all measurements $\mathbf{y} = \{y_i(s), i = 1, \ldots, r; s = 1, \ldots, n\}$ with $\boldsymbol{\beta} = \{\boldsymbol{\beta}(s), s = 1, \ldots, n\}$ which represents a high-dimensional vector of space-varying coefficients. This approach considers not only spatial correlation within the model but also improves the tensor fitting by the diffusivity information of adjacent voxels. Moreover, the use of spline-basis functions serves to model the spatial field $\mathbf{D} = \{\mathbf{D}(\mathbf{s}), \mathbf{s} \in \mathbb{R}^3\}$, i. e. the field $\boldsymbol{\beta} = \{\boldsymbol{\beta}(\mathbf{s}), \mathbf{s} \in \mathbb{R}^3\}$. As a consequence, estimates for $\mathbf{D}(\mathbf{s})$ are accessible at any arbitrary position $\mathbf{s} \in \mathbb{R}^3$ in the brain, not only on the discrete grid of voxels $s = 1, \ldots, n$. This approximation property is an important prerequisite of any DTI tracking algorithm which needs estimates of $\mathbf{D}(\mathbf{s})$ on a markedly finer 3d grid than available by the acquisition resolution in order to reconstruct fiber bundles in a biologically smooth fashion. Current techniques use either some sort of simple interpolation technique between voxels in a subsequent (third) step (Gössl et al., 2002) or aim at a combination of smoothing and interpolation (Pajevic et al., 2002; Mishra et al., 2006). In a probablistic framework, the interpolated data might be taken from a single neighboring point that has the highest probability functional of its position relative to the interpolation site (Behrens et al., 2003).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

At first glance, our proposed concept of a space-varying coefficient model seems to be straightforward. However, the massive dimension of the 3d array $\boldsymbol{\beta} = \{\boldsymbol{\beta}(s), s = 1, \ldots, n\}$ of coefficient vectors of length six, thus of a 4d array, implies methodological and computational challenges which cannot be solved adequately with existing methods. Originally, VCMs have been suggested by Hastie and Tibshirani (1993) for regressions with coefficients varying smoothly over a one-dimensional continuous variable such as time-varying effects. Extensions to 2d-space-varying coefficients have been developed more recently, ranging from 2d-surface smoothers to (Markov) random field models (Assunção, 2003; Eilers and Marx, 2003; Fahrmeir et al., 2004). Further extensions from two dimensions to three- or higher-dimensional arrays are more challenging than it may seem at first sight and much less effort has been spent on this topic, with the notable exception of the array regression method proposed by Eilers et al. (2006). As suggested in Heim et al. (2004), array regression was first employed to analyze DTI data with a VCM model in Eilers et al. (2005), but unfortunately does not allow for enough knots in 3d-space. This is a severe limitation for diffusion imaging analysis as important anatomical details need to be retained.

3.2 Diffusion tensor imaging

DTI is named according to its fundamental quantity, the field $\{\mathbf{D}(\mathbf{s}), \mathbf{s} \in \mathbb{R}^3\}$ of diffusion tensors

$$\mathbf{D}(\mathbf{s}) = \begin{pmatrix} D_1(\mathbf{s}) & D_4(\mathbf{s}) & D_5(\mathbf{s}) \\ D_4(\mathbf{s}) & D_2(\mathbf{s}) & D_6(\mathbf{s}) \\ D_5(\mathbf{s}) & D_6(\mathbf{s}) & D_3(\mathbf{s}) \end{pmatrix}.$$

A diffusion tensor characterizes the local diffusivity of water molecules at each location $\mathbf{s} \in \mathbb{S} \subset \mathbb{R}^3$ within a continuous subspace $\mathbb{S} \subset \mathbb{R}^3$ of the brain. From a mathematical point of view, the water molecules follow a random walk in space, conceptualized as (an-)isotropic 3d Brownian motion with location dependent diffusion matrix $\mathbf{D}(\mathbf{s})$. A probabilistic model for the physical diffusion of particles is given by the diffusion process $\mathrm{X} = \{\mathrm{X}(t), 0 \leq t_0 \leq t \leq T\}$ with state space $\mathbb{S} \subset \mathbb{R}^3$ (Grimmett and Stirzaker, 2001). The reading is that a particle, e. g. a water molecule, has the spatial coordinate $\mathrm{X}(t)$ at

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

time t. Technically, the stochastic differential equation

$$d\mathrm{X}(t) = \mathbf{D}^{\frac{1}{2}}(\mathrm{X}(t))d\mathrm{W}(t) \tag{3.2}$$

holds, where the diffusion process is driven by the 3d continuous-time Brownian process $\mathrm{W} = \{\mathrm{W}(t), t \geq 0\}$, also called Wiener process. The interpretation is that the infinitesimal location shift $d\mathrm{X}(t)$ depends on an infinitesimal error term weighted with the diffusion matrix. Under the assumption of normality, the error term can be replaced by the random shock of a 3d White-Noise process which results from the small difference $d\mathrm{W}(t)$. The location and time dependent diffusion matrix determines the distribution of the random error and, thereby, how much and in which directions random movements are possible. Hence, the tensor itself can be interpreted as the (local) covariance matrix of an (anisotropic) Wiener process. The eigenvalues $\xi_i(\mathbf{s}), i = 1, 2, 3$, and eigenvectors of this symmetric and (theoretically) positive definite matrix correspond to the axis lengths and directions of the local diffusion ellipsoid. Among the eigenvalue based intravoxel measures, fractional anisotropy (FA) is the most popular metric to assess the local degree of anisotropy of the diffusion process. By definition (Basser and Pierpaoli, 1996)

$$\mathrm{FA}(\mathbf{s}) = \sqrt{\frac{3}{2}\frac{\sum_i(\xi_i(\mathbf{s}) - \bar{\xi}(\mathbf{s}))^2}{\sum_i \xi_i^2(\mathbf{s})}}$$

with mean eigenvalue $\bar{\xi} = \frac{1}{3}\sum_{i=1}^{3}\xi_i$, the FA index is zero in a perfect isotropic medium and equals one in a purely anisotropic medium. FA and other eigenvalue based metrics are used to characterize the physiological microstructure of brain tissue and neuropathological processes such as stroke (Sotak, 2002), brain tumours or inflammatory disease (Horsfield and Jones, 2002). However, both the analysis of specified brain regions and whole brain approaches such as voxel based morphometry (see Mechelli et al. (2005) and the references therein) are hampered by the generally coarse spatial resolution of diffusion images and by error prolongation caused by noise. Therefore, sufficient approximation techniques for high resolution mapping of small scale processes and complex anatomical structures are desirable.

On the other hand, the dominant eigenvector of the diffusion ellipsoid reflects the principal diffusion direction which is particularly pronounced in white matter: Neural fibers in this brain compartment are densely packed and highly ordered such that the water

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

molecules therein preferentially pass along the biophysiological structures instead of perpendicularly to them (Basser and Jones, 2002). Hence the main eigenvector is assumed to (approximately) coincide with the local fiber orientation in space. This postulate underlies all existing tracking algorithms which ultimately aim at reconstructing neural fiber bundles (Mori and van Zijl, 2002; Le Bihan, 2003). Figure 3.1 outlines this basic idea for a certain slice of a 3d data grid. The gray curve corresponds to a stylized part of a

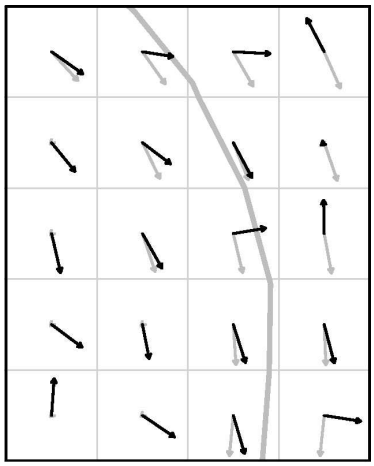


Figure 3.1: Vector map of template (gray) and recorded (black) principal diffusivity directions; the underlying fiber tract is indicated as gray curve. Note, at the central fiber-transit voxel, the severe distortion of the main eigenvector due to noise.

fiber bundle, and for each discrete grid point, indexed by s, the gray arrow indicates the dominant eigenvector of the "true" unobservable diffusion tensor $\mathbf{D}(s)$. The orientation of these true eigenvectors varies more or less randomly in isotropic parts of the volume while eigenvectors associated with the fiber bundle correspond to anisotropic diffusion tensors and provide good information about local directionality of the neural tract, that is still to be reconstructed. Linking together the eigenvectors by some suitable tracking algorithm allows to recover the neural tract. However, as described below, only (noisy) tensor estimates $\hat{\mathbf{D}}(s)$ can be obtained from the recorded DTI data and hence the derived principal eigenvectors will likely be erroneous as exemplified by black arrows in Fig. 3.1. Clearly, tracking algorithms which work on with these unsatisfactory principal eigenvectors can yield spurious results. See, for instance, the severely distorted main eigenvector in the middle fiber-transit voxel of Fig. 3.1. Thus, there is an evident need to denoise raw estimates by some kind of spatial smoothing, borrowing strength from adjacent voxels.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

In the following we describe the standard approach of obtaining estimates of $\mathbf{D}(\mathbf{s})$ as well as corresponding eigenvalues and -vectors from DTI raw data.

Physical theory states that the relation between the magnetic resonance hardware and the recorded images is given by the (deterministic) Stejskal-Tanner equation (Stejskal and Tanner, 1965):

$$S_i(s) \;=\; S_0(s)\exp\left\{-b\,\mathbf{g}_i{}'\mathbf{D}(s)\mathbf{g}_i\right\}\,, \quad i=1,\ldots,r\,, \quad s=1,\ldots,n. \tag{3.3}$$

Here $S_0(s)$ is the non-diffusion weighted reference intensity at voxel s, while $S_i(s)$ is the signal intensity in the same voxel measured after the application of the ith diffusion weighting magnetic resonance gradient $\mathbf{g}_i = (g_{1i}, g_{2i}, g_{3i})'$ (see Mori and Barker (1999) for physical background). The scalar b comprises several acquisition parameters such as gradient strength and duration; $\mathbf{D}(s)$ denotes the local diffusion tensor. Yet in practice, the relation (3.3) is disturbed by thermal and physiological noise. Following Papadakis et al. (1999), Eq. (3.3) can be reformulated, for voxel s, $s = 1,\ldots,n$, as

$$y_i(s) = -\frac{1}{b}\log\left(\frac{S_i(s)}{S_0(s)}\right) \;=\; \mathbf{x}_i{}'\boldsymbol{\beta}(s) + \varepsilon_i(s), \quad i=1,\ldots,r, \tag{3.4}$$

with noise $\varepsilon_i(s) \overset{\text{iid}}{\sim} N(0,\sigma^2)$, the vector $\boldsymbol{\beta}(s) = (D_1, D_2, D_3, D_4, D_5, D_6)'(s)$ of unknown elements of the diffusion tensor and the design vector $\mathbf{x}_i = (g_{1i}^2, g_{2i}^2, g_{3i}^2, 2g_{1i}g_{2i}, 2g_{1i}g_{3i}, 2g_{2i}g_{3i})'$, constructed from the known values of $\mathbf{g}_i$. In Eq. (3.4) we recognize a linear regression problem with the unknown six-dimensional vector $\boldsymbol{\beta}(s)$. Note that $\mathbf{x}_i$ is the same for all voxels, and thus we face a (spatially) repeated measurement design with $(r\times 6)$-dimensional covariate matrix $\mathbf{X} = (\mathbf{x}_1,\ldots,\mathbf{x}_r)'$. The space invariance of $\mathbf{X}$ plays a role for efficient implementation (see Section 3.8). To give an example, a typical gradient scheme implies the following covariate design:

$$\mathbf{X} \;=\; \begin{pmatrix} 0.5 & 0.5 & 0.0 & 0.0 & 0.5 & 0.5 \\ 0.0 & 0.0 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & -1.0 \\ 1.0 & -1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & -1.0 & 0.0 & 0.0 \end{pmatrix}.$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

For the pupose of a solution to Eq. (3.4), at least $r \geq 6$ diffusion weighted images from non-collinear magnetic gradient directions are required. Usually, measurements are repeated for the same gradient set or the number of different gradients is chosen to be overdetermined in order to increase the signal-to-noise ratio. Due to ethical as well as economic reasons, the time frame for clinical imaging protocols is limited, requiring optimal off-line denoising algorithms.

Presumably, the simplifying assumption of independently and identically distributed noise in Eq. (3.4) looks quite restrictive, and it might seem desirable to allow for temporally or spatially correlated errors $\varepsilon_i(s), i = 1, \ldots, r, s = 1, \ldots, n$. Note, however, that model (3.4) assumes the spatial variation of the observed signals $y_i(s)$ to be already taken into account through spatially varying coefficients $\boldsymbol{\beta}(s)$, thereby interpreting $\varepsilon_i(s)$ as pure measurement noise without any serial dependence. Allowing for spatially heterogeneous variances $\sigma^2(s) = \mathrm{Var}(\varepsilon_i(s))$ is possible, but will still lead to the same separate OLS estimates $\hat{\boldsymbol{\beta}}(s)$. In the standard approach, spatial correlation is considered implicitly in a postprocessing step through spatial smoothing of the estimated field $\{\hat{\boldsymbol{\beta}}(s), s = 1, \ldots, n\}$, for example by applying Gaussian kernels or other spatial smoothers. The space-varying coefficient model introduced in the next section automatically accounts for spatial smoothing and correlation right in the model formulation.

3.3 The space-varying coefficient model

Our novel concept combines the n separate regression models of the standard approach to a joint space-varying coefficient model by a suitable (spatial) design of the 3d array $\{\boldsymbol{\beta}(s), s = 1, \ldots, n\}$ of coefficient vectors $\boldsymbol{\beta}(s) = (\beta_1, \ldots, \beta_p)'(s)$:

$$\mathbf{y} = \sum_{j=1}^{p} \mathbf{Z}_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_{rn}), \tag{3.5}$$

where the observations in the rn-dimensional vector $\mathbf{y}$ are ordered according to repeats first and voxel number last and $\boldsymbol{\beta}_j = (\beta_j(1), \ldots, \beta_j(n))'$ denotes the vectorized jth coefficient surface. The $(rn \times n)$-matrix $\mathbf{Z}_j = \mathbf{I}_n \otimes \mathbf{X}(\cdot, j)$ with $(n \times n)$-identity matrix $\mathbf{I}_n$ consists of n blocks, each one containing the r-dimensional regressor $\mathbf{X}(\cdot, j)$. Thus, spatial correlation and information from adjacent voxels are taken into account. In addition, the

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

number of repetitions, r, can be kept small as it is desirable to avoid long acquisition times. An intercept, though not required by the specific application of DTI with $p = 6$, can be incorporated in model (3.5) by adding a term $\mathbf{Z}_0\boldsymbol{\beta}_0$, where $\mathbf{Z}_0 = (1, \ldots, 1)' \in \mathbb{R}^{rn}$.

3.3.1 Multidimensional B-splines

Aiming at a flexible non-parametric representation of the 3d varying coefficient surfaces $\{\beta_j(s), s = 1, \ldots, n\}$, $j = 1, \ldots, p$, a linear basis expansion in s is a popular device. This section introduces the employed basis functions, i. e. the B(asic)-spline (curves) family. A comprehensive introduction on the theory and properties of (penalized) B-splines can be found in Dierckx (1993), Hämmerlin and Hoffmann (1989) and Eilers and Marx (1996).

In general, a linear basis expansion in x,

$$f(x) = \sum_{k=1}^{K} \gamma_k B_k(x),$$

is an appropriate device to achieve more flexible representations for $f(x)$, where the basis function $B_k(x) : \mathbb{R}^p \to \mathbb{R}$ is the kth transformation of x and γ_k the corresponding weight. Such basis functions are typically determined by smoothness (e. g. the degree of the B-spline) and by position (the number and placement of knots).

In one dimension, a B-spline basis function $B_k(x)$ of degree g and with knots $\kappa_k, \ldots, \kappa_{k+g+1}$ is defined through the following recurrence relation:

$$\begin{aligned} B_k^g(x) &= \frac{x - \kappa_k}{\kappa_{k+g} - \kappa_k} B_k^{g-1}(x) + \frac{\kappa_{k+g+1} - x}{\kappa_{k+g+1} - \kappa_{k+1}} B_{k+1}^{g-1}(x), \\ B_k^0(x) &= \begin{cases} 1 & x \in [\kappa_k, \kappa_{k+1}) \\ 0 & x \notin [\kappa_k, \kappa_{k+1}). \end{cases} \end{aligned}$$

Thus, a B-spline extends over $g + 2$ knots and is nonzero in the corresponding interval. Zero-degree B-splines correspond to piecewise constants. Since their use does not enable interpolation, they are not suited for the estimation of the diffusion tensor field with the ultimate purpose of fiber tracking. In Fig. 3.2 as well as in the present application, an equidistant knot setting is applied. While the derivative of linear basis functions is zero at the knots, higher order B-splines are smoother and in shape similar to Gaussian kernels.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Usually, the basis dimension is large, so that instruments are required for controlling the complexity of the model. For example, regularization can be achieved through squared difference penalties on the coefficients γ_k (Eilers and Marx, 1996). This leads to the penalized maximum likelihood estimate for $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)'$:

$$\hat{\boldsymbol{\gamma}} \quad = \quad (\mathbf{B}_1'\mathbf{B}_1 + \lambda_1 \boldsymbol{\Delta}_1'\boldsymbol{\Delta}_1)^{-1}\mathbf{B}_1'\mathbf{y},$$

where $\mathbf{y}$ is the n_1-dimensional observation vector, the $(n_1 \times K)$-matrix $\mathbf{B}_1$ contains the basis functions evaluated at the observation points, and $\boldsymbol{\Delta}_1$ denotes the $(K-d\times K)$-matrix of dth order difference penalties. The smoothing parameter λ_1 controls the trade-off between data closeness and parsimony in the model variables. An example of a data-driven penalized fit of the amplitudes is given in Fig. 3.2. Obviously, a smooth function (gray curve) should be reconstructed using smooth basis functions, i. e. of higher degree. Increasing the tuning parameter λ_1 is not a remedy here, since this results in flattened curvatures of the reproduced function.

The extension to two dimensions $(x, y) \in \mathbb{R}^2$ is achieved by building the $(K \times L)$-dimensional tensor product basis:

$$B_{kl}^g(x, y) \quad = \quad B_k^g(x)B_l^g(y), \quad k = 1, \ldots, K; l = 1, \ldots, L.$$

Representatives of different degree are depicted in Fig. 3.3. The contour plots reveal clear anisotropy in case of bilinear B-splines. This behaviour caused by knot-discontinuity is

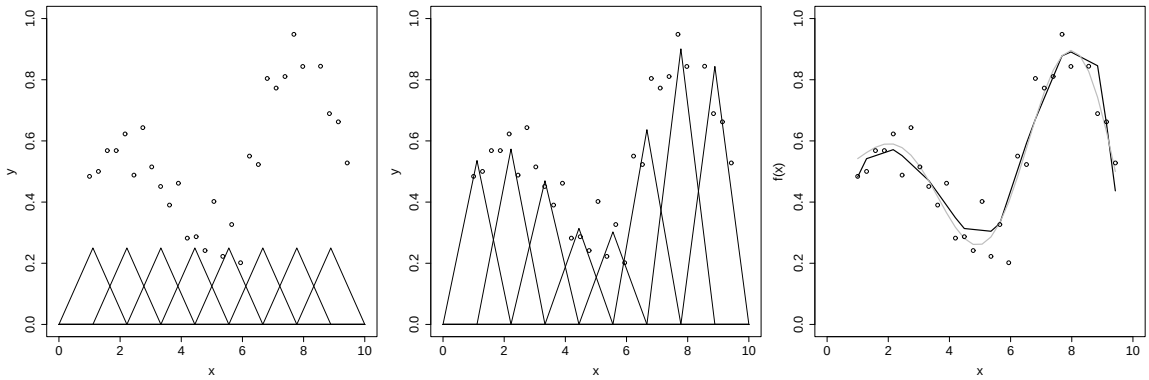


Figure 3.2: On the left side, noisy data is to be fitted by a non-parametric function using linear B-spline basis functions. In the center, the basis functions are scaled according to their estimated weights. On the right side, the resulting fit (black) and the true function (gray) are superimposed to the noisy data.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

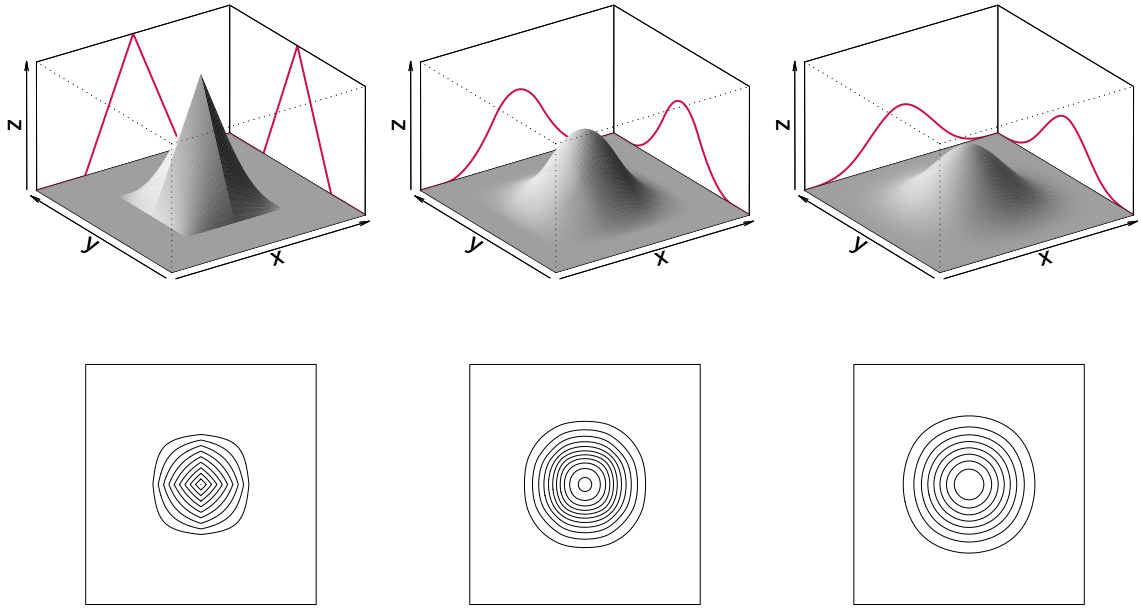


Figure 3.3: Surface and contour plots of 2d B-spline basis functions of linear, quadratic, and cubic degree (left to right).

attenuated but still present for the quadratic case, and only from cubic degree on hardly visible to the human eye. In principle, a dimension specific degree is conceivable.

A 2d surface spline can consequently be represented by a linear combination of basis functions $B^g_{kl}(x, y)$. Figure 3.4 gives a schematic impression of the linear expansion in (x, y).

Generalization to further dimensions is straightforward but lacks graphical visualization. Penalization is particularly indispensible in higher dimensions as the number of basis functions grows exponentially. Thereby it holds that the effective number of degrees of freedom and, thus, the flexibility of the penalized spline falls between the penalty order and the number of basis functions.

3.3.2 Multidimensional smoothing with tensor products

The elements of $\boldsymbol{\beta}_j = \{\beta_j(s), s = 1, \ldots, n\}, j = 1, \ldots, p$, are modelled non-parametrically by projecting them onto penalized tensor product B-splines or multidimensional P-splines.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

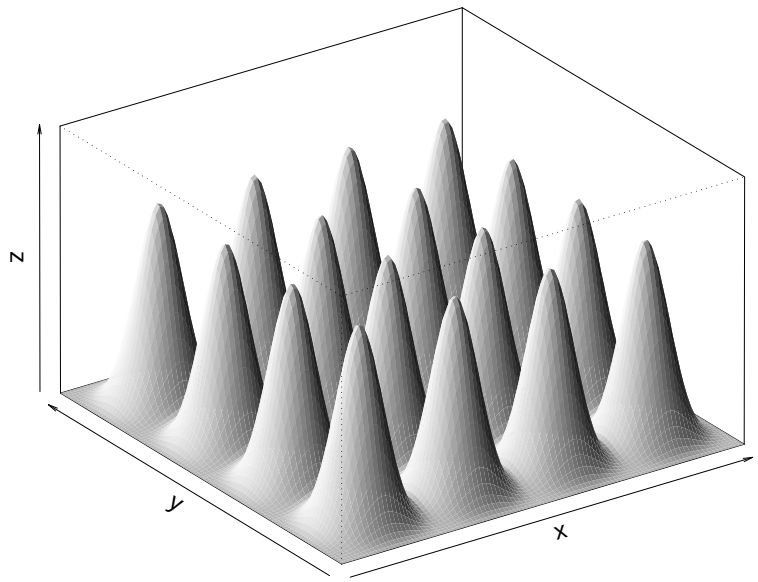


Figure 3.4: A selection of cubic 2d B-splines which, in a full basis, are more densely packed with considerable overlap.

We first give a direct VCM presentation. Consider

$$\boldsymbol{\beta}_j(s) = \sum_{v=1}^{KLM} \mathbf{B}(s,v)\gamma_j(v) = \mathbf{B}(s,\cdot)\boldsymbol{\gamma}_j,$$

where the $(n \times KLM)$-matrix $\mathbf{B} = \mathbf{B}_3 \otimes \mathbf{B}_2 \otimes \mathbf{B}_1$ contains the 3d tensor products of 1d B-splines, i. e. of $(n_1 \times K)$-matrix $\mathbf{B}_1$, $(n_2 \times L)$-matrix $\mathbf{B}_2$, and $(n_3 \times M)$-matrix $\mathbf{B}_3$ evaluated at x-, y-, and z-coordinates, respectively. Regarding the full jth coefficient surface we equivalently write

$$\boldsymbol{\beta}_j = \mathbf{B}\boldsymbol{\gamma}_j. \tag{3.6}$$

Note that $\mathbf{B}$ is the same for all coefficient surfaces $\boldsymbol{\beta}_j$ and has to be calculated only once. The vector $\boldsymbol{\gamma}_j$ denotes the unknown amplitudes of the basis functions, and $K \times L \times M$ is determined by the (generous and regularly gridded) knot partition and the degree of the basis functions. To ensure sufficiently fine reconstruction of essential features in DTI, one knot is required at each 1.25 voxel as suggested by a pilot study. Furthermore, the spline degree is set linear in order to keep the influencing regions of the basis functions as locally restricted as possible. There is nothing prohibative in the general methodology from using other choices of basis degree.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Using the tensor coefficient expression in (3.6), the least squares criterion derived from Eq. (3.5) becomes

$$LS(\boldsymbol{\gamma}) = \left\| \mathbf{y} - \sum_{j=1}^{p} \mathbf{Z}_j \boldsymbol{\beta}_j \right\|^2 = \left\| \mathbf{y} - \sum_{j=1}^{p} \mathbf{Z}_j \mathbf{B} \boldsymbol{\gamma}_j \right\|^2 ,$$

where $\mathbf{B}$ is of dimension $n \times KLM$, $\boldsymbol{\gamma}_j$ of $KLM \times 1$, and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_p')'$. We then aim to find a practical solution to the penalized objective

$$\begin{aligned} LS_{pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}) &= LS(\boldsymbol{\gamma}) + \mathrm{Pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}) \\ &= \|\mathbf{y} - (\mathbf{B} \otimes \mathbf{X})\boldsymbol{\gamma}\|^2 + \mathrm{Pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}). \end{aligned} \tag{3.7}$$

Note that the second equality of $LS(\boldsymbol{\gamma})$ in (3.7) holds because the p interaction variables are space-invariant in the present application of DT imaging and, hence, the n blocks in $\mathbf{Z}_j$ are the same. This restatement plays an important role in the implementation.
The penalty term $\mathrm{Pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma})$ serves to avoid overfitting and is tuned by the parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)' \in \mathbb{R}_+^3$. For each regressor j, difference penalties are placed on the rows, columns, and layers of tensor product coefficients, such that

$$\begin{aligned} \mathrm{Pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}) &= \sum_{j=1}^{p} \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{m=1}^{M} \left\{ \lambda_1 (\Delta_1^d \gamma_{jklm})^2 + \lambda_2 (\Delta_2^d \gamma_{jklm})^2 + \lambda_3 (\Delta_3^d \gamma_{jklm})^2 \right\} \\ &= \lambda_1 \|(\mathbf{I}_L \otimes \mathbf{I}_M \otimes \boldsymbol{\Delta}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma}\|^2 + \lambda_2 \|(\mathbf{I}_M \otimes \boldsymbol{\Delta}_2 \otimes \mathbf{I}_K \otimes \mathbf{I}_p)\boldsymbol{\gamma}\|^2 \\ &\quad + \lambda_3 \|(\boldsymbol{\Delta}_3 \otimes \mathbf{I}_K \otimes \mathbf{I}_L \otimes \mathbf{I}_p)\boldsymbol{\gamma}\|^2 \\ &= \lambda_1 \|\mathbf{P}_1 \boldsymbol{\gamma}\|^2 + \lambda_2 \|\mathbf{P}_2 \boldsymbol{\gamma}\|^2 + \lambda_3 \|\mathbf{P}_3 \boldsymbol{\gamma}\|^2 . \end{aligned} \tag{3.8}$$

Here Δ_1^d, Δ_2^d, and Δ_3^d denote the d-th order differences across a row, down a column, and along a layer of the $K \times L \times M$ array of tensor product B-spline coefficients, $\boldsymbol{\Gamma}_j = [\gamma^j{}_{klm}]$, respectively. Note that, in general, different values of d are conceivable for the three penalized dimensions. In the present application to DTI data, $d = 1$ proved most appropriate with respect to satisfactory detail sustainment. The $(K - d \times K)$-matrix $\boldsymbol{\Delta}_1$, $(L - d \times L)$-matrix $\boldsymbol{\Delta}_2$, and $(M - d \times M)$-matrix $\boldsymbol{\Delta}_3$ indicate the corresponding matrices of difference penalties, while $\mathbf{P}_1 = (\mathbf{I}_L \otimes \mathbf{I}_M \otimes \boldsymbol{\Delta}_1 \otimes \mathbf{I}_p)$, $\mathbf{P}_2 = (\mathbf{I}_M \otimes \boldsymbol{\Delta}_2 \otimes \mathbf{I}_K \otimes \mathbf{I}_p)$, and $\mathbf{P}_3 = (\boldsymbol{\Delta}_3 \otimes \mathbf{I}_K \otimes \mathbf{I}_L \otimes \mathbf{I}_p)$ consists of a carefully arranged, full matrix representation using block diagonal matrices of right Kronecker products. If $\lambda_1 \neq \lambda_2 \neq \lambda_3$, the smoothing parameter is dimension specific. In contrast, the smoothing parameter is global if $\lambda_1 =$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

$\lambda_2 = \lambda_3$. In both cases it determines the trade-off between smoothness and fidelity to the data. Details on data-driven optimization of the tuning parameter are given in Section 3.8.

In theory, an explicit solution to $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_p')'$ can be found using

$$\hat{\boldsymbol{\gamma}} = (\mathbf{U}'\mathbf{U} + \mathbf{P})^{-1}\mathbf{U}'\mathbf{y}, \tag{3.9}$$

where $\mathbf{U} = g([\mathbf{Z}_1\mathbf{B}, \ldots, \mathbf{Z}_p\mathbf{B}]) = \mathbf{B} \otimes \mathbf{X}$ is $(rn \times pKLM)$-dimensional, and $\mathbf{P} = \lambda_1\mathbf{P}_1'\mathbf{P}_1 + \lambda_2\mathbf{P}_2'\mathbf{P}_2 + \lambda_3\mathbf{P}_3'\mathbf{P}_3$ is derived from (3.8). The auxiliary function g serves to reorder the columns of its argument so that the ordering rule "repeats first, voxel numbers last" is met. Given $\boldsymbol{\gamma}$, then the varying coefficient volumes can be built. Since the basis functions can routinely be evaluated at a large number of (intermediate) points, the resolution can be increased straightforwardly without applying an additional interpolation method.

For practical brain imaging applications, we may need about $K \times L \times M = 32 \times 32 \times 8$ knots for a region of interest sized $40 \times 40 \times 10$ voxels. Lowering the number of knots means a wider support for the basis functions and, in consequence, a tendence for smearing essential features as investigated in a pilot study. Introducing substantially more knots permits noise to sneak in the estimation result. Thus $\mathbf{B}$ has approximately 3×10^9 elements: if each floating point takes 8 bytes, then $\mathbf{B}$ will use several Gb of memory, which is beyond reach of current computers. The left hand side of the normal equations preceding Eq. (3.9) would occupy more than 18 Gb for the above number of parameters, i. e. $6 \times 32 \times 32 \times 8$. Hence efficient algorithmic implementation and programming using sparsity and approximations is essential (see also Section 3.8).

In principle, the SVCM (Eq. (3.5)) as well as the penalized least squares criterion (Eq. (3.7)) can be extended to account for error variances $\varepsilon_i(s)$ varying across voxels by introducing diagonal weight matrices $\mathbf{W} = \mathrm{diag}(\mathrm{W}(s), s = 1, \ldots, n)$ and $\mathrm{W}(s) = \mathrm{diag}(\sigma_i^2(s), i = 1, \ldots, r)$ and modifying the least squares criterion $LS(\boldsymbol{\gamma})$ accordingly. Estimation could be carried out in two steps: First estimate coefficients through penalized LS based on Eq. (3.7), and variances $\sigma^2(s)$ at each voxel separately through averaged residual sums of squares. Second, re-estimate coefficients through minimization of the weighted version of Eq. (3.7). Note, however, that at least a large part of spatial variation is already covered through the space-varying coefficients, so that the additional compu-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

tational effort compared to improved statistical efficiency becomes rather questionable. Therefore, we keep to the simplifying assumption of spatially homogeneous measurement error variances. Also bear in mind that smoothing does consider possible correlation among the observations. This becomes particularly evident from the Bayesian perspective, where penalization arises when $\{\boldsymbol{\beta}(s), s = 1, \ldots, n\}$ is considered to be random and endowed with Markov random field priors (see Brezger and Lang, 2006).

3.3.3 Successive smoothing with univariate basis functions

A simplification of the SVCM version with tensor products can be derived from Dierckx (1982) and Dierckx (1993, p. 172) who proposes the so-called "new smoothing spline" for penalized 2d smoothing. The key point of the procedure concerns the data arrangement which is left in multidimensional array structure. For demonstration purpose, the procedure is presented in 2d and without any regressor matrix, i. e. for the case of simple smoothing of 2d data. Dierckx searches for a solution to the equation system given by

$$\mathbf{y} = (\mathbf{B}_2 \otimes \mathbf{B}_1)\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I}) \tag{3.10}$$

with constraints

$$\sqrt{\lambda_1}(\mathbf{B}_2 \otimes \boldsymbol{\Delta}_1)\boldsymbol{\gamma} = \mathbf{0}_{(K-d)n_2 \times 1} \tag{3.11}$$

$$\sqrt{\lambda_2}(\boldsymbol{\Delta}_2 \otimes \mathbf{B}_1)\boldsymbol{\gamma} = \mathbf{0}_{(L-d)n_1 \times 1} \tag{3.12}$$

$$\sqrt{\lambda_1\lambda_2}(\boldsymbol{\Delta}_2 \otimes \boldsymbol{\Delta}_1)\boldsymbol{\gamma} = \mathbf{0}_{(K-d)(L-d) \times 1}. \tag{3.13}$$

Starting from the univariate B-spline matrices augmented by the corresponding difference penalties,

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{B}_1 \\ \sqrt{\lambda_1}\boldsymbol{\Delta}_1 \end{pmatrix}, \qquad \mathbf{A}_2 = \begin{pmatrix} \mathbf{B}_2 \\ \sqrt{\lambda_2}\boldsymbol{\Delta}_2 \end{pmatrix},$$

and

$$\mathbf{y}_+ \;=\; \mathrm{vec}(\mathbf{Y}_+) = \mathrm{vec}\begin{pmatrix} \mathbf{Y}_{n_1 \times n_2} & \mathbf{0}_{n_1 \times (L-d)} \\ \mathbf{0}_{(K-d) \times n_2} & \mathbf{0}_{(K-d) \times (L-d)} \end{pmatrix}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

as well as $\boldsymbol{\gamma} = \text{vec}(\boldsymbol{\Gamma}_{K\times L})$, the objective corresponding to Eqs. (3.10) - (3.13) can be reformulated:

$$\begin{aligned} LS_{pen}(\boldsymbol{\lambda},\boldsymbol{\gamma}) &= LS(\boldsymbol{\gamma}) + \text{Pen}(\boldsymbol{\lambda},\boldsymbol{\gamma}) \\ &= \|\mathbf{y} - (\mathbf{B}_2 \otimes \mathbf{B}_1)\boldsymbol{\gamma}\|^2 + \\ &\quad \left\|\sqrt{\lambda_1}(\mathbf{B}_2 \otimes \boldsymbol{\Delta}_1)\boldsymbol{\gamma}\right\|^2 + \left\|\sqrt{\lambda_2}(\boldsymbol{\Delta}_2 \otimes \mathbf{B}_1)\boldsymbol{\gamma}\right\|^2 + \left\|\sqrt{\lambda_1\lambda_2}(\boldsymbol{\Delta}_2 \otimes \boldsymbol{\Delta}_1)\boldsymbol{\gamma}\right\|^2 \\ &= \left\|\mathbf{y}_+ - (\mathbf{A}_2 \otimes \mathbf{A}_1)\boldsymbol{\gamma}\right\|^2. \end{aligned}$$

Recall substantial properties of the Kronecker product as, for example, quoted in Dierckx (1993, p. 170):

$$(\mathbf{V} \otimes \mathbf{W})' = \mathbf{V}' \otimes \mathbf{W}' \tag{3.14}$$

$$(\mathbf{VW}) \otimes (\mathbf{CD}) = (\mathbf{V} \otimes \mathbf{C})(\mathbf{W} \otimes \mathbf{D}) \tag{3.15}$$

$$\text{vec}(\mathbf{V}_{p\times n}\, \mathbf{X}_{n\times m}\, \mathbf{W}_{m\times q}) = (\mathbf{W}' \otimes \mathbf{V})\,\text{vec}(\mathbf{X}), \tag{3.16}$$

where the vec-operator performs column-wise stacking of its argument. By the use of data augmentation and the above properties, the respective normal equations can be transformed as follows:

$$\begin{aligned} & (\mathbf{A}_2 \otimes \mathbf{A}_1)'(\mathbf{A}_2 \otimes \mathbf{A}_1)\,\text{vec}(\boldsymbol{\Gamma}) = (\mathbf{A}_2 \otimes \mathbf{A}_1)'\,\text{vec}(\mathbf{Y}_+) \\ \overset{(3.14),(3.15)}{\Longleftrightarrow} \quad & (\mathbf{A}_2'\mathbf{A}_2 \otimes \mathbf{A}_1'\mathbf{A}_1)\,\text{vec}(\boldsymbol{\Gamma}) = (\mathbf{A}_2' \otimes \mathbf{A}_1')\,\text{vec}(\mathbf{Y}_+) \\ \overset{(3.16)}{\Longleftrightarrow} \quad & \text{vec}(\mathbf{A}_1'\mathbf{A}_1\boldsymbol{\Gamma}\mathbf{A}_2'\mathbf{A}_2) = \text{vec}(\mathbf{A}_1'\mathbf{Y}_+\mathbf{A}_2) \\ \Longleftrightarrow \quad & (\mathbf{B}_1'\mathbf{B}_1 + \lambda_1\boldsymbol{\Delta}_1'\boldsymbol{\Delta}_1)\boldsymbol{\Gamma}(\mathbf{B}_2'\mathbf{B}_2 + \lambda_2\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2) = \mathbf{B}_1'\mathbf{Y}\mathbf{B}_2 \end{aligned}$$

$$\Longrightarrow \quad \boldsymbol{\Gamma} = (\mathbf{B}_1'\mathbf{B}_1 + \lambda_1\boldsymbol{\Delta}_1'\boldsymbol{\Delta}_1)^{-1}\mathbf{B}_1' \underbrace{\mathbf{Y}\mathbf{B}_2(\mathbf{B}_2'\mathbf{B}_2 + \lambda_2\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2)^{-1}}_{=\boldsymbol{\Gamma}_{\text{tmp}}',\ (n_1\times L)\text{-matrix}}$$

$$\text{with} \quad \boldsymbol{\Gamma}_{\text{tmp}} = (\mathbf{B}_2'\mathbf{B}_2 + \lambda_2\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2)^{-1}\mathbf{B}_2'\mathbf{Y}'.$$

Apparently, an estimate for the coefficients $\boldsymbol{\Gamma}$ results from consecutive univariate smoothing with the observational data as input to the smoothing cycle depending on $\mathbf{B}_2$. The (preliminary) coefficients $\boldsymbol{\Gamma}_{\text{tmp}}$ serve each as input to further iterations, here along the other dimension. Insertion into the model Eq. (3.10) yields the fitted values and gives

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

insight into the structure of the so-called hat matrix (= smoother matrix):

$$\begin{aligned}
\hat{\mathbf{y}} &= (\mathbf{B}_2 \otimes \mathbf{B}_1)\,\mathrm{vec}(\hat{\mathbf{\Gamma}}) \\
&\overset{(3.16)}{=} (\mathbf{B}_2 \otimes \mathbf{B}_1)\left(\left[(\mathbf{B}_2'\mathbf{B}_2 + \lambda_2\mathbf{\Delta}_2'\mathbf{\Delta}_2)^{-1}\mathbf{B}_2'\right] \otimes \left[(\mathbf{B}_1'\mathbf{B}_1 + \lambda_1\mathbf{\Delta}_1'\mathbf{\Delta}_1)^{-1}\mathbf{B}_1'\right]\right)\mathrm{vec}(\mathbf{Y}) \\
&\overset{(3.15)}{=} \underbrace{\left[\mathbf{B}_2(\mathbf{B}_2'\mathbf{B}_2 + \lambda_2\mathbf{\Delta}_2'\mathbf{\Delta}_2)^{-1}\mathbf{B}_2'\right] \otimes \left[\mathbf{B}_1(\mathbf{B}_1'\mathbf{B}_1 + \lambda_1\mathbf{\Delta}_1'\mathbf{\Delta}_1)^{-1}\mathbf{B}_1'\right]}_{=\mathbf{H},\ \text{hat matrix}}\mathbf{y}
\end{aligned}$$

Generalization to three dimensions including the regressor matrix $\mathbf{X}$ leads to an seemingly unpleasant equation system:

$$\mathbf{y} = (\mathbf{B}_3 \otimes \mathbf{B}_2 \otimes \mathbf{B}_1 \otimes \mathbf{X})\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I})$$

with constraints

$$\begin{aligned}
(\mathbf{B}_3 \otimes \mathbf{B}_2 \otimes \mathbf{B}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_1}(\mathbf{B}_3 \otimes \mathbf{B}_2 \otimes \mathbf{\Delta}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_2}(\mathbf{B}_3 \otimes \mathbf{\Delta}_2 \otimes \mathbf{B}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_3}(\mathbf{\Delta}_3 \otimes \mathbf{B}_2 \otimes \mathbf{B}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_1\lambda_2}(\mathbf{B}_3 \otimes \mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_1\lambda_3}(\mathbf{\Delta}_3 \otimes \mathbf{B}_2 \otimes \mathbf{\Delta}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_2\lambda_3}(\mathbf{\Delta}_3 \otimes \mathbf{\Delta}_2 \otimes \mathbf{B}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_1\lambda_2\lambda_3}(\mathbf{\Delta}_3 \otimes \mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1 \otimes \mathbf{I}_p)\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_1}(\mathbf{B}_3 \otimes \mathbf{B}_2 \otimes \mathbf{\Delta}_1 \otimes \mathbf{X})\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_2}(\mathbf{B}_3 \otimes \mathbf{\Delta}_2 \otimes \mathbf{B}_1 \otimes \mathbf{X})\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_3}(\mathbf{\Delta}_3 \otimes \mathbf{B}_2 \otimes \mathbf{B}_1 \otimes \mathbf{X})\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_1\lambda_2}(\mathbf{B}_3 \otimes \mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1 \otimes \mathbf{X})\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_1\lambda_3}(\mathbf{\Delta}_3 \otimes \mathbf{B}_2 \otimes \mathbf{\Delta}_1 \otimes \mathbf{X})\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_2\lambda_3}(\mathbf{\Delta}_3 \otimes \mathbf{\Delta}_2 \otimes \mathbf{B}_1 \otimes \mathbf{X})\boldsymbol{\gamma} &= \mathbf{0} \\
\sqrt{\lambda_1\lambda_2\lambda_3}(\mathbf{\Delta}_3 \otimes \mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1 \otimes \mathbf{X})\boldsymbol{\gamma} &= \mathbf{0}.
\end{aligned}$$

As before, the use of augmented matrices

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{X} \\ \mathbf{I}_p \end{pmatrix} \qquad \mathbf{A}_i = \begin{pmatrix} \mathbf{B}_i \\ \sqrt{\lambda_i}\mathbf{\Delta}_i \end{pmatrix}, \quad i = 1, 2, 3.$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

allows for a simplified notation of the objective function

$$LS_{pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}) \quad = \quad \|\mathbf{y}_+ - (\mathbf{A}_3 \otimes \mathbf{A}_2 \otimes \mathbf{A}_1 \otimes \mathbf{A}_0)\boldsymbol{\gamma}\|^2$$

and the corresponding hat matrix

$$\mathbf{H} \quad = \quad \mathbf{B}_3(\mathbf{A}_3'\mathbf{A}_3)^{-1}\mathbf{B}_3' \otimes \mathbf{B}_2(\mathbf{A}_2'\mathbf{A}_2)^{-1}\mathbf{B}_2' \otimes \mathbf{B}_1(\mathbf{A}_1'\mathbf{A}_1)^{-1}\mathbf{B}_1' \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \qquad (3.17)$$

Though hidden by the augmented notation, Dierckx's objective function emcompasses a set of Kronecker product terms: one purely consisting of design matrices, one purely consisting of penalty matrices and the remaining terms being mixtures of all possible combinations. The number of penalty terms increases in *power of dimension* with Dierckx's approach, but only in *dimension* with the tensor product approach defined by Eqs. (3.7) and (3.8). Hence, the sequential smoothing approximates the tensor product approach. To elaborate the difference, consider augmentation in the tensor product case:

$$\mathbf{A} = \begin{pmatrix} \mathbf{B}_3 \otimes \mathbf{B}_2 \otimes \mathbf{B}_1 \otimes \mathbf{X} \\ \sqrt{\lambda_1}(\mathbf{I}_{LM} \otimes \boldsymbol{\Delta}_1 \otimes \mathbf{I}_p) \\ \sqrt{\lambda_2}(\mathbf{I}_M \otimes \boldsymbol{\Delta}_2 \otimes \mathbf{I}_{Kp}) \\ \sqrt{\lambda_3}(\boldsymbol{\Delta}_3 \otimes \mathbf{I}_{KLp}) \end{pmatrix} = \begin{pmatrix} \mathbf{U} \\ \sqrt{\lambda_1}\mathbf{P}_1 \\ \sqrt{\lambda_2}\mathbf{P}_2 \\ \sqrt{\lambda_3}\mathbf{P}_3 \end{pmatrix}, \qquad \mathbf{y}_+ = \begin{pmatrix} \mathbf{y} \\ \mathbf{0}_{p(K-d)LM} \\ \mathbf{0}_{pK(L-d)M} \\ \mathbf{0}_{pKL(M-d)} \end{pmatrix},$$

with subsequent restatement of Eq. (3.7):

$$LS_{pen}(\boldsymbol{\lambda}, \boldsymbol{\gamma}) \quad = \quad \|\mathbf{y}_+ - \mathbf{A}\boldsymbol{\gamma}\|^2.$$

The resulting hat matrix takes on the form

$$\begin{aligned} \mathbf{H}^{TP} &= \mathbf{U}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{U}' \\ &= \mathbf{U}\left(\mathbf{U}'\mathbf{U} + \lambda_1\mathbf{P}_1'\mathbf{P}_1 + \lambda_2\mathbf{P}_2'\mathbf{P}_2 + \lambda_3\mathbf{P}_3'\mathbf{P}_3\right)^{-1}\mathbf{U}' \\ &= \mathbf{U}\left(\mathbf{U}'\mathbf{U} + \mathbf{P}\right)^{-1}\mathbf{U}'. \end{aligned}$$

The generalization to higher dimensions can be implemented on the basis of the function $\rho(\mathbf{V}, \mathbf{W}, i)$ published by Eilers et al. (2006) and described in detail by Currie et al. (2006). The argument $\mathbf{V}$ is a 2d matrix, $\mathbf{W}$ a higher-dimensional array, and i a dimension index. This function performs the usual computations of the matrix product, along the rows of $\mathbf{V}$ and along dimension i of $\mathbf{W}$. The core trick is to rotate dimension i of $\mathbf{W}$ to the front

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

by dimension permutation, reduce the array $\mathbf{W}$ to two dimensions, compute the standard matrix product with $\mathbf{V}$, transform $\mathbf{W}$ back to its original dimensionality and rotate the ith dimension back to place. Repeated application of ρ to the signal intensities $\mathbf{Y}$ yields the following $(p \times K \times L \times M)$-array estimate of $\mathbf{\Gamma} = \mathrm{vec}(\boldsymbol{\gamma})$:

$$\begin{aligned}\hat{\mathbf{\Gamma}} &= \rho\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',\ \rho\left((\mathbf{B}_1'\mathbf{B}_1 + \lambda_1\boldsymbol{\Delta}_1'\boldsymbol{\Delta}_1)^{-1}\mathbf{B}_1',\ \rho\left((\mathbf{B}_2'\mathbf{B}_2 + \lambda_2\boldsymbol{\Delta}_2'\boldsymbol{\Delta}_2)^{-1}\mathbf{B}_2',\right.\right.\right.\\ &\quad \left.\left.\left.\rho\left((\mathbf{B}_3'\mathbf{B}_3 + \lambda_3\boldsymbol{\Delta}_3'\boldsymbol{\Delta}_3)^{-1}\mathbf{B}_3',\ \mathbf{Y},\ 4\right),\ 3\right),\ 2\right),\ 1\right). \qquad (3.18)\end{aligned}$$

Starting from the inner brackets, ρ smoothes the fourth dimension of $\mathbf{Y}$, thereby transforming the original $(r \times n_1 \times n_2 \times n_3)$-dimensional $\mathbf{Y}$ to a $(r \times n_1 \times n_2 \times M)$-array. Analogous proceeding renders this $(r \times n_1 \times n_2 \times M)$-array $(r \times n_1 \times L \times M)$-dimensional, then $(r \times K \times L \times M)$- and finally $(p \times K \times L \times M)$-dimensional. Note that sequential univariate smoothing according to Eq. (3.18) is highly attractive in terms of working memory and computation time. This approach affords an entirely different opportunity to overcome implementational deficiencies. Due to the sequential character of the procedure, involved matrices are small a priori allowing to place a large number of basis functions. Moreover, the crucial part of the generalized cross-validation (GCV), namely the trace of the hat matrix, can be computed very efficiently since the trace of Kronecker products equals the product of traces:

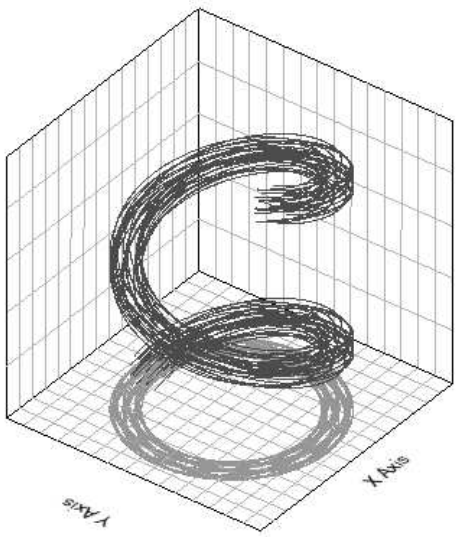
$$\mathrm{tr}(\mathbf{H}) = \left(\sum_{i=1}^{r} \mathrm{diag}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(i,i)\right)\prod_{j=1}^{3}\left(\sum_{i=n_1}^{n_j}\mathbf{H}_j(i,i)\right), \qquad (3.19)$$

with $\mathbf{H}_j = \mathbf{B}_j(\mathbf{B}_j'\mathbf{B}_j + \lambda_j\boldsymbol{\Delta}_j'\boldsymbol{\Delta}_j)^{-1}\mathbf{B}_j'$. For further speed-up the trace arguments can be rotated.

3.4 Simulation study

For the purpose of performance rating of the SVCM approaches versus the standard estimation method, a simulation study is conducted with quasi realistic parameter settings and a fiber tract following the geometry of a spiral. Due to the lack of a gold standard, the spiral is promoted as a conventional simulation model within DTI literature. Figure 3.5 displays the stylized fiber bundle together with the auxillary grid of anisotropic voxel size typical to clinical experiments. Since each fiber-transit voxel is considered a fiber

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.



- 3d data grid of $\{15 \times 15 \times 5\} \subset \mathbb{R}^3$ voxels
- $2 \times 2 \times 4$ mm^3 voxel size
- six 3d varying coefficient surfaces
- spiral tensors are anisotropic; background tensors are isotropic
- simulated Gaussian error with $\sigma = 10$

Figure 3.5: Design of the simulation study and geometry of the underlying fiber bundle.

voxel (compare e. g. the projection on the xy-plane), the overall ratio of fiber to non-fiber voxels amounts to 0.2.

In the given simulation model, the background tensors are spherically shaped corresponding to a totally isotropic diffusion process. Yet, in spiral voxels the tensor shape is cigar-like. This means that the underlying diffusion process is unambiguously one-directional. In other terms, the template eigenvalues determining the spiral tensors obey a ratio of 2:1:1 giving a fractional anisotropy value of FA = 0.4. We defer to Fig. 3.8 for a complete map of the true tensors. The imposed amount of Gaussian error is taken from 70 real data sets as the average background noise. Also the voxel size is chosen to mimic the specifications generally used in a DTI experiment.

In the following, ST1 denotes the standard procedure of voxelwise regression using ordinary least squares methodology. The analysis technique is abbreviated ST2 if smoothing with a Gaussian kernel of full width half maximum[1] equalling 0.75 voxels is appended, and ST3 for additional trilinear interpolation on top of the estimated and regularized data. Among the SVCM variants, realization with tensor product basis functions, as well as successive univariate smoothing according to Dierckx, are examined constrained to global and dimension specific smoothing parameters, respectively. We will refer to these

[1] In the case of Gaussian kernel smoothing, the full width half maximum denotes an alternative measure for the width of the Gaussian shape and is related to the scale parameter σ by $FWHM = 2\sqrt{2\log 2}\sigma$.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

approaches as TPglob, TP3d, DXglob, and DX3d. To ensure sufficient reconstruction of fine, although essential features, one knot is required at each 1.25 voxels in both artificial and real data. Linear spline basis functions and first order difference penalties further promote texture preservation. The total number of regression coefficients, i. e. of amplitudes of the basis functions, amounts to $12 \times 12 \times 4 \times 6 = 3456$ in all SVCM approaches. Note that the spectral decomposition yields least erroneous eigenvalues and eigenvectors if tensor elements are estimated most accurately. Therefore the analysis is restricted to the fitted coefficient surfaces $\{\hat{\boldsymbol{\beta}}_j, j = 1, \ldots, 6\}$ constituting the tensor field, and to the derived 3d field of FA which represents the most widely used scalar measure of anisotropy.

Concerning the performance of the proposed methods, TPglob (TP3d) took on average about 19 seconds with fixed tuning parameter and 10.46 minutes (21.08 minutes) if the smoothing parameter was optimized by GCV. DXglob (DX3d) needed 1 seconds and 3 seconds (4 seconds), respectively. In case of the reference procedures, ST1 took 1 seconds, subsequent Gaussian smoothing 8 seconds and trilinear interpolation 1 seconds. According to this, ST3 needed about 10 seconds. The time quoted refers to a 2.4 GHz Pentium IV with 1 Gb memory. These results are essentially reflecting the computational complexity of each method, as described in Section 3. Note that the performance strongly depends on the data volume size and behaves non-linearly, in particular in the tensor product approach. Clearly, the choice of the starting point affects the time until the (local) minimum is reached.

For each of the $N = 100$ simulation runs, we assessed the overall quality of the competing estimation procedures by the averaged mean squared error (AMSE) of the tensor fit, depending on the voxel type 'spiral':

$$\mathrm{AMSE}_{\mathrm{sp}}^{(i)} = \frac{1}{p\, n_{\mathrm{sp}}} \sum_{s \in \mathrm{spiral}} \left\| \boldsymbol{\beta}(s) - \hat{\boldsymbol{\beta}}^{(i)}(s) \right\|^2, \quad i = 1, \ldots, N,$$

or 'background':

$$\mathrm{AMSE}_{\mathrm{bg}}^{(i)} = \frac{1}{p\, n_{\mathrm{bg}}} \sum_{s \in \mathrm{background}} \left\| \boldsymbol{\beta}(s) - \hat{\boldsymbol{\beta}}^{(i)}(s) \right\|^2, \quad i = 1, \ldots, N.$$

The $\boldsymbol{\beta}(s)$ and $\hat{\boldsymbol{\beta}}^{(i)}(s)$ denote the template coefficient vector and its counterpart fitted from the ith simulation run, respectively. Note that AMSE is a region specific error measure

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

per simulation run, averaged over all region-related voxels and all tensor elements. Analogous definitions hold for FA.

Thus, we can obtain empirical error distributions of the estimated coefficient surfaces and also of the FA values once confined to fiber tissue (spiral) and once confined to background tissue. From the comparison of the relevant boxplots, we are able to judge which estimation method performs best with regard to tissue types with different diffusivity properties.

With respect to the spiral voxels that are of particular importance for fiber tracking, Figure 3.6 confirms the superiority of the SVCM approaches over the current state-of-the-art ST2 when fitting the tensor components. DX3d performs best, followed by DXglob, TP3d, TPglob and ST2 with medians at −18.98, −18.93, −18.85, −18.83, and −18.53 (left boxplots of top subfigure). This ranking changes to general advantage of 3d tuning parameter when the resolution is duplicated as evident from the left boxplots of the bottom subfigure. On this last level of data processing prior to fiber tracking, the various SVCM approaches outperform the ST3 procedure (median = −18.24) with the clearest

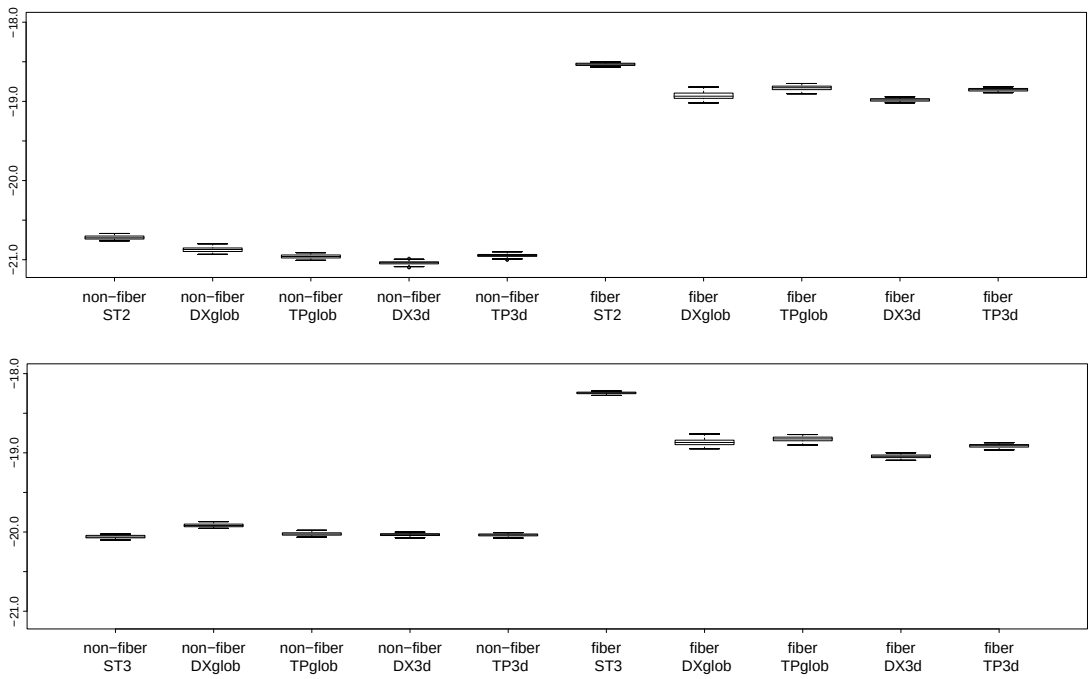


Figure 3.6: Log AMSE of all six tensor elements estimated by different approaches and grouped according to background (left) and spiral voxels (right). Boxplots comprise the voxel volume at original (top) and duplicated (bottom) resolution.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

improvement yielded by DX3d (-19.04), followed by TP3d (-18.91), DXglob (-18.87) and lastly by TPglob (-18.82). In the isotropic background, error distributions appear rather similar between the considered approaches both at original and refined resolution. If the fitted tensor field is transformed to the 3d FA field, almost the same relationship of quality of fit holds for the different techniques: FA is fitted more accurately in the background than in the spiral voxels, where the various SVCM approaches lead to clearly smaller errors than ST2 (cf., Fig. 3.7). This latter distinction becomes more pronounced with respect to interpolated estimates though the ranking remains exactly the same, i. e. DX3d, TP3d, TPglob, DXglob, and ST2/ST3 in decreasing performance order. It is worth mentioning that the SVCM variants result in even smaller errors of spiral than background FA estimates, if the degree of anisotropy is further augmented to an eigenvalue ratio of 10:1:1 (FA = 0.89; data not shown). In the background compartment, all methods show a tendency of larger errors at increased resolution compared to original acquisition.

Concerning AMSE of both tensor and FA estimation, SVCM approaches with a 3d penalty perform altogether better than those with a global tuning parameter. Based on these results and from a computational point of view, we clearly favor DX3d. For comparison, TP3d and ST2 are also examined in more detail, namely for the tensor estimation.

To glean intravoxel quality information, we compute for each coefficient surface $\boldsymbol{\beta}_j, j = 1, \ldots, 6$, the voxelwise mean squared error (VMSE) defined as

$$\mathrm{VMSE}_j(s) \quad = \quad \frac{1}{N} \sum_{i=1}^{N} \left(\boldsymbol{\beta}_j(s) - \hat{\boldsymbol{\beta}}_j^{(i)}(s) \right)^2, \quad s = 1, \ldots, n.$$

This measure reflects the error of each coefficient at each voxel s, averaged over the simulation runs.

Figure 3.8 displays the error ratio of two respective estimation procedures on a log scale. The VSME map corresponding to one method is set relative to the VMSE map resulting from the other method. Taking the logarithm leads to a symmetric scale of the ratio. For example, if $\log(\mathrm{VMSE}_{\mathrm{DX3d}}/\mathrm{VMSE}_{\mathrm{ST2}}) = 2$ this corresponds to an error ratio of $\exp(2) = 7.389 \approx 7 : 1$, meaning that method DX3d results in seven times larger errors than the standard ST2.

In general, this type of graphic allows to assess both the relative magnitude and the

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

distributional structure of the errors. The emerging color-coded pattern suggests that the standard method mainly results in more biased tensor components *at the ridges* while the basis function approaches lead to larger errors *at the edges* (see bottom row of Fig. 3.8). This becomes obvious from comparison of the log error ratios (bottom row) with the template structure (top left). Whereas green spots occur above all at true edges and in a second instance in the background, rose colored points tend to build up artificial ridges which can be associated with spiral echoes. For example, the second slice from below shows a u-shaped semicircle for all true tensor elements. Exactly at the edges of these semicircles, the SVCM approach performs worse than the standard procedure (green points). Reversely, the standard procedure exhibits larger errors than the SVCM approach not only at the true ridges, but also in the upper half of the slice (rose points). Yet compared to DX3d (bottom left), the TP3d variant (bottom right) seems to lack successful reproduction of the spiral structure: The upper half of the second slice contains slightly more rim artifacts on the basis of TP3d than of DX3d (see green spots).

Note that errors of the dark-green range do hardly occur at all whereas we can detect a

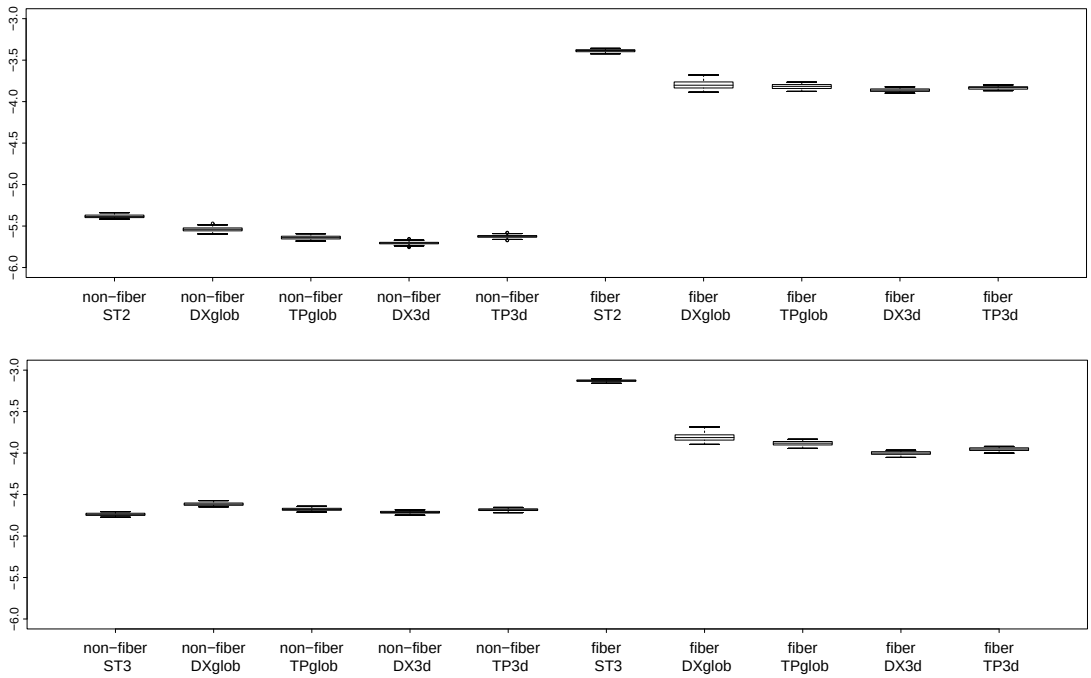


Figure 3.7: Log AMSE of FA values for different estimation approaches based on the original (top) and the duplicated resolution (bottom). Distinction refers to background (left) and spiral voxels (right).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

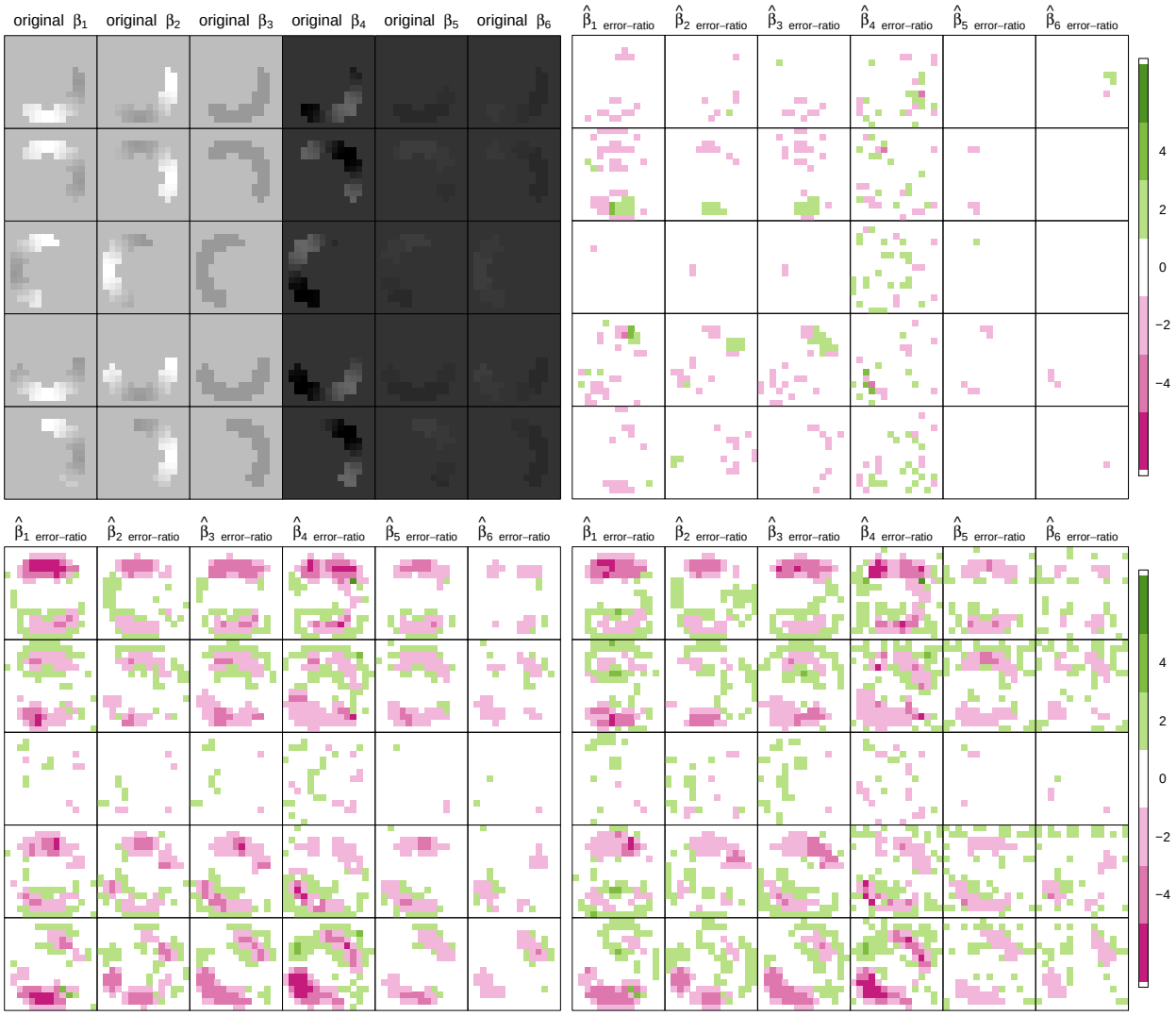


Figure 3.8: Template coefficient surfaces (top left) corresponding to the six elements of the 3d tensor field. Log ratio of VMSE is given for both the DX3d (bottom left) and the TP3d approach (bottom right) relative to ST2. Top right shows the log ratio of VMSE from DX3d versus TP3d.

considerable number of points colored dark-rose. With respect to DX3d (bottom left), the impression arises that the same proportional shift holds for the medium saturated nuances. This can be understood as superiority of SVCM approaches over the standard procedure.

Analogous comparison of the two best SVCM approaches against each other is included top right in Fig. 3.8. The prevailing texture of green and rose colored voxels suggests that TP3d does not agglomerate as large errors in artificial clusters as the DX3d method (green), but loses in accuracy at true edges (rose).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

To conclude, there is an obvious preferability of the SVCM approaches compared to the current standard, in particular, if the advantage of inherent approximation and, thus, resolution refinement is considered.

3.4.1 Supplement

A supplementary simulation study was conducted with respect to different acquistion schemes, including either three times repeated application of the same set of diffusion weighting MR gradients (3 [repeats] $\times$ 6 [gradients]) or a set of more than six gradients (1×15, 1×31). All other parameters were left untouched as in Fig. 3.5. For comparability, results from the preceding study design (1×6) are also considered.
To assess the impact of the acquistion scheme, the comparison focuses on the methods of voxelwise regression (ST1), voxelwise regression plus Gaussian kernel smoothing (ST2), voxelwise regression plus Gaussian kernel smoothing plus trilinear interpolation (ST3) and the sequential variant of the SVCM with first order difference penalties, global tuning parameter, linear (DXglob-lin) and quadratic (DXglob-quad) B-spline basis functions as well as 1 knot per 1.25 voxels. The choice of a 3d smoothing parameter had negligible effect.

AMSE at original resolution: Independent of the acquisition scheme, both DXglob-lin and DXglob-quad outperform standard ST2 but not ST1. This benefit is most distinct in anisotropic voxels. DXglob-quad performs slightly better than DXglob-lin. Interestingly, more repeats do, in general, not result in the desired effect of improved tensor estimation (e. g. with ST1). In contrast, a larger number of gradients leads to smaller errors, in particular when tensors are estimated using ST1 or DXglob-quad. In summary, ST1-(1×31) performs clearly best.

AMSE at double resolution: In anisotropic conditions, also both DXglob-lin and DXglob-quad are superior to ST3, however, both visual inspection and MSE analyses show background inhomogeneity. Neither increasing directions nor repeats could improve SVCM estimation quality over all voxels. Importantly, interpolation using DXglob-lin outperformed trilinearly interpolated ST1 or ST3 for all acquisition schemes.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

VMSE log ratio at original resolution: While ST2 tends to blur the spiral and to evoke additional shape artifacts, the SVCM approaches catches the "height" better but on the cost of inferior background smoothness and overshooting at edges, in particular with DXglob-quad. This pattern points to inadequate edge preservation. If less gradients are used, less Gibbs phenomena appear. Clearly, DXglob-lin at 1×6 gradients leads to equal or even better estimation quality than ST2 throughout the whole volume apart from small and isolated Gibbs artifacts.

VMSE log ratio at double resolution: VMSE log ratio maps expose the same method specific deficiencies of either shape and intensity distortions (standard procedures) or wiggly background reproduction in combination with Gibbs phenomena (SVCM).

3.5 Application to real data

Besides from being the preferential estimation procedure according to the simulation study, successive smoothing with dimension specific penalties exhibits several computational advantages for application to real data (see Discussion and Section 3.8). The data set was shrunken from a recorded number of 128×128 within plane voxels to 90×75 due to non-informative background voxels. Furthermore a selection of six subsequent slices was considered sufficient for fiber tracking of, for instance, the visual system. From the remaining volume of human brain images, the diffusion tensor field was derived using ST1, ST2 and DX3d. The same knot density as in the simulation study served to maintain important biological features. In total, $72 \times 60 \times 5 \times 6$ coefficients needed to be determined in the SVCM; this are $113,400$ unknown parameters less than in the standard model of voxelwise regression.

Figure 3.9 depicts the respective results of one exemplary diagonal tensor element, i. e. $\hat{\boldsymbol{\beta}}_{2,\mathrm{ST1}}$, $\hat{\boldsymbol{\beta}}_{2,\mathrm{ST2}}$ and $\hat{\boldsymbol{\beta}}_{2,\mathrm{DX3d}}$ with an inferior-superior ordering of the axial slices from left to right. When the focus is on fiber tracking, standard estimation on the basis of voxelwise regression (ST1, top row) is usually regularized to eliminate noise artifacts as in the present manner of Gössl et al. (2002) (ST2, middle row). The bottom slices depict the estimated surface achieved by the SVCM application. Looking at the second and sixth

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

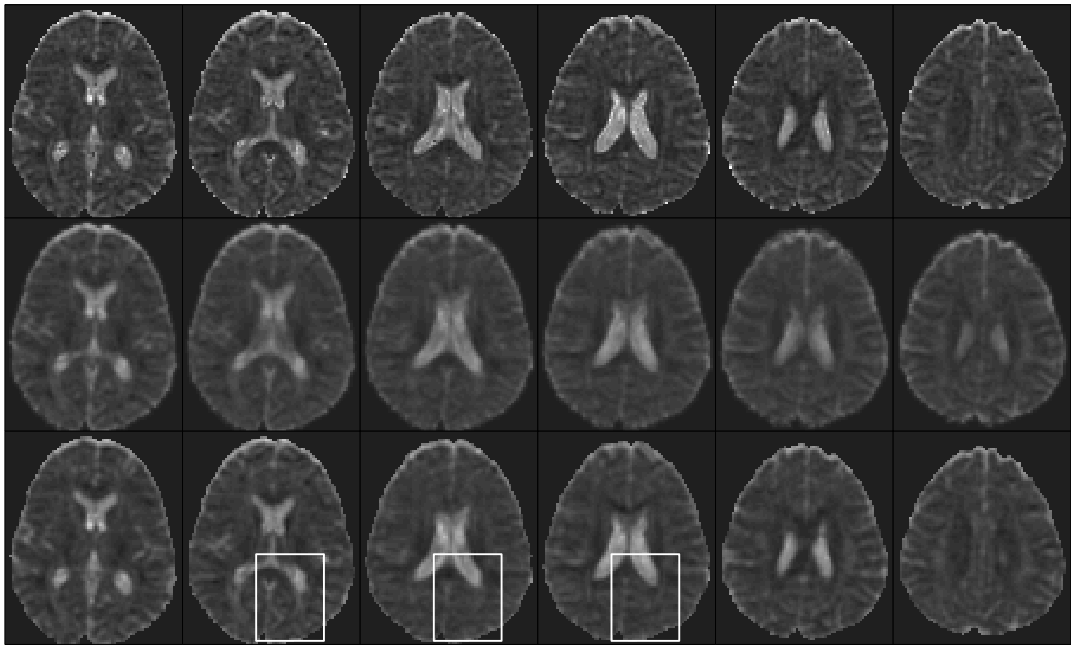


Figure 3.9: Exemplary diagonal element of the diffusion tensor estimated by mere voxelwise regression (top row), plus subsequent regularization with a Gaussian kernel (middle row) and by applying the SVCM with DX3d (bottom row). The white rectangulars indicate the segment that was interpolated to double resolution as shown in Fig. 3.10.

column, in particular, reveals an obvious oversmoothing of the Gaussian kernel causing ventricles (bright blobs) to be severely smeared across neighbouring slices.

The described algorithm which unifies estimation, regularization and interpolation can enhance typical image post-processing steps in brain mapping as image segmentation and coregistration. Especially non-linear spatial transformation steps needed for the matching of interindividual brain anatomy can benefit from a continuous vector field. Refinement of the observational grid – as also required by each tracking algorithm – is demonstrated for a smaller area superimposed as white sketch (Fig. 3.9). The box segment comprises three tissue types (cerebro-spinal fluid, gray and white matter) holding distinct diffusivity properties. In practice, some interpolation method leading to ST3 is appended to the standard 2-step procedure in order to allow for quasi continuous fiber reconstruction. With respect to the alternative SVCM, additional evaluation of the basis functions at intermediate points complies with this demand. Figure 3.10 delineates the close-up of the respective segment layers including two intermediate planes at double resolution: The SVCM estimation (bottom) reveals by far more details and achieves a stronger contrast

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

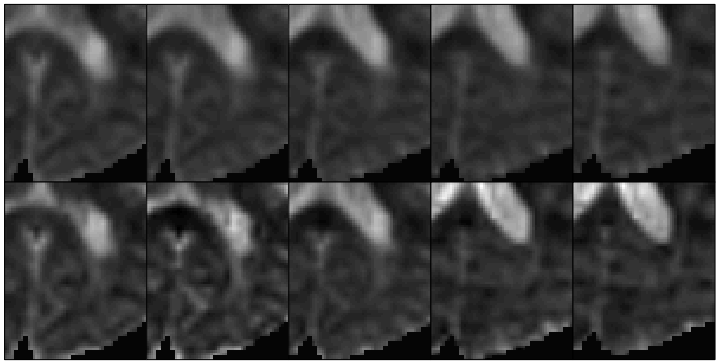


Figure 3.10: Top row represents the exemplary diagonal element of the diffusion tensor at duplicated resolution when the current 3-step data processing is applied. Juxtaposed are the corresponding results yielded by the SVCM model (bottom row).

than the standard data preprocessing cascade ST3 (here with trilinear interpolation, top row).

3.6 Discussion

In this chapter, we proposed a space-varying coefficient model with 3d basis functions as an alternative estimation method in diffusion tensor imaging. Potential regularization and interpolation steps are covered within the same framework without error propagation and accumulation. This makes up a salient property of the SVCM approach with basis functions, namely the inherent possibility of going beyond the rigid grid resolution limited by the data acquisition process. A simulation study proved the superiority of our SVCM variants in terms of averaged mean squared error (AMSE) and voxelwise mean squared error (VMSE) to the current standard procedure of voxelwise regression with subsequent regularization. For real data a higher contrast of the diffusion tensor field could be achieved than with trilinear interpolation that completes the standard estimation method. We solely succeeded to handle the clinical example of $90 \times 75 \times 6$ selected voxels by applying the sequential and thus computationally advantageous DX3d VCM. Despite the approximate character, it had also turned out to be the 'candidate of choice' in the simulation study. In contrast to a global penalization, the optimization of a dimension specific tuning parameter showed to account for the commonly anisotropic voxel size more appropriately as evident from slightly smaller mean squared errors.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Linear B-splines were used in all cases although cubic B-splines are conventionally preferred for their quasi-isotropy. For DT application however, the sustainment of important biological features is of primary concern, but likely to be impaired by too large influence regions of adjacent voxels. Therefore, a lower spline degree seems more appropriate as confirmed from a pilot study. First order difference penalties and a knot at every 1.25 voxel were additionally chosen for both simulated and real data. Unfortunately, the demands on computer memory and computation time load increase sharply with knot density. We faced this challenge by the extensive use of the sparsity of the involved matrices, hence solving the penalized normal equations in an efficient way again. Yet, for 3d images of the original size it will essentially be impossible to explicitly form the necessary tensor products. The sequential methods DXglob and DX3d represent so far the one and only possibility to handle such big problems. See Section 3.8 for remarks on implementational matters.

Despite all efforts to preserve enough detail, the proposed SVCM approach still suffers from so-called Gibbs phenomena as obvious from Fig. 3.8 of the VMSE quotient. This under- and overshoot around discontinuities or areas with high curvature appears when the reconstruction of a discontinuous or rapidly changing function is aimed for by a set of continuous ones. Similar phenomena are also known from the one- and two-dimensional case, i. e. in non-parametric curve and surface estimation, and various proposals have been made to improve local adaptivity of estimators. Two main concepts seem promising: The first one still relies on spline basis functions but introduces spatially adaptive penalties, see e. g. (Ruppert and Carroll, 2000; Lang et al., 2002; Brezger et al., 2005) for the 1d case, and (Brezger et al., 2005) for the 2d case. The latter paper shows that computational demands already grow enormously when extending 1d-methods to two dimensions. A second possibility is to switch to different basis functions such as wavelets and radial basis functions, or to base the model on anisotropic random fields. Though being conceptually published in 2d, all these locally adaptive modifications are again extremely challenging in 3d from an algorithmic point of view.

Regarding possible methodological generalizations, our approach naturally accommodates observation weights, so missing data does not represent a problem and adjustments can be made for voxels with uncertainty, e. g. at the boundaries. Furthermore, as parameter

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

estimation is grounded in classical methods, model summaries, such as effective dimension, information criteria, and delete-one diagnostics are available. Confidence intervals can be obtained with the iterative method, but their accessibility can be of practical matter with the direct tensor product approach depending on the total number of coefficients. Last but not least, our penalized 3d space-varying VCM approach can be transplanted into a generalized linear model framework, e. g. smoothing coefficients associated with a binary or threshold response. This way, also non-linearity of the Stejskal-Tanner equation could be considered.

We demonstrated the superiority of the proposed model with respect to estimation and interpolation of the diffusion tensor field. As a direct benefit, secondary measures of anisotropy (e. g. FA) or diffusivity can be calculated at higher resolution to assess complex brain regions such as the hippocampus. Since there is still a discrepancy between imaging resolution and microscopical fiber tract organization of the brain, fiber tracking at high-resolution is warranted in particular due to its growing role for neurosurgical operation planning and generation of probability atlas systems. Yet, prior to a daily use of the proposed SVCM method, an extended simulation study is required to corroborate the performance in more general scenarios. Such a study should at least encompass different choices of tensor shapes, anisotropy degrees and noise levels. More precise neuronal tract reconstruction is expected but has to be proven, e. g. on the basis of geometrically more elaborated phantom images[2] and suitable validation measures such as the average angle deviation from the template vector field or an appropriate similarity measure (Mishra et al., 2006). The linkage of the SVCM with a respective algorithm has not been accomplished and constitutes an on-going research topic.

3.7 Explanatory remarks

As pointed out in the discussion (Section 3.6), the SVCM approach could be improved with respect to occuring detrimental effects. In the following, these two drawbacks are further illustrated.

[2] http://neurology.iop.kcl.ac.uk/dtidataset/Common_DTI_Dataset.htm by Derek Jones and Sean Deoni

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3.7.1 Gibbs artifacts

Ringing artifacts are also known as Gibbs phenomena that denote the typical behaviour of Fourier series around discontinuities.

Let $f(x), x \in \mathbb{R}$ be a square wave function with period $P = 2l$:

$$f(x) \quad = \quad f(x \pm (2l)k) = \begin{cases} -1 & x \in [-l, 0) \\ 1 & x \in [0, l) \end{cases}, k \in \mathbb{Z}.$$

Then the Fourier representation

$$f(x) \quad = \quad a_0 + \sum_{j=1}^{\infty} \left(a_j \cos\left(\frac{2\pi jx}{P}\right) + b_j \sin\left(\frac{2\pi jx}{P}\right)\right)$$

with coefficients

$$a_j = \frac{1}{l} \int_{-l}^{l} f(x) \cos\left(\frac{2\pi jx}{P}\right) dx, j = 0, 1, \ldots$$

and

$$b_j = \frac{1}{l} \int_{-l}^{l} f(x) \sin\left(\frac{2\pi jx}{P}\right) dx, j = 1, \ldots$$

resolves to

$$f(x) \quad = \quad \sum_{j=1}^{\infty} \left(\frac{4}{\pi(2j-1)} \sin\left(\frac{2\pi(2j-1)x}{P}\right)\right). \tag{3.20}$$

The cosine terms all vanish in Eq. (3.20) because the a_j's are inner products of the *odd* impulse function, i. e. $f(x) = -f(-x)$, with each of the *even* cosine functions and, hence, have zero weight. The coefficients of the sine terms are derived as follows:

$$\begin{aligned} b_j &= \frac{2}{P} \int_{-l}^{l} f(x) \sin\left(\frac{2\pi jx}{P}\right) dx \\ &= \frac{2}{P} \left(\int_{-l}^{0} -\sin\left(\frac{2\pi jx}{P}\right) dx + \int_{0}^{l} \sin\left(\frac{2\pi jx}{P}\right) dx \right) \\ &= \frac{2}{P} \left(\left[\frac{P}{2\pi j} \cos\left(\frac{2\pi jx}{P}\right)\right]_{x=-l}^{x=0} - \left[\frac{P}{2\pi j} \cos\left(\frac{2\pi jx}{P}\right)\right]_{x=0}^{x=l} \right) \\ &= \frac{2}{P} \frac{P}{2\pi j} \left(\cos(0) - \cos(-\pi j) - \cos(\pi j) + \cos(0)\right) \\ &= \frac{4}{\pi j} \end{aligned}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

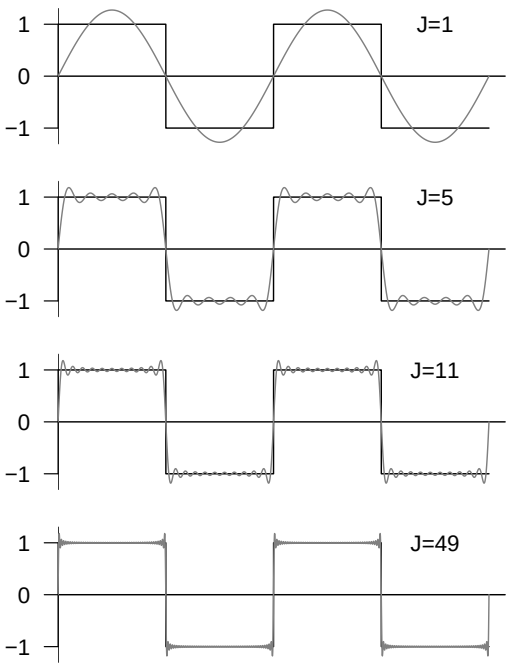


Figure 3.11: Approximation of a discontinuous signal by Fourier expansion with J terms.

Figure 3.11 depicts a rectangular impulse function in black with its Fourier expansion superimposed in gray. With increasing number J of cosine-sine terms, the approximation quality improves except for the jumps. This is due to the pointwise convergence of the Fourier series at discontinuities. As a consequence, the typical over- and undershoots arise. Transferred to the brain data, different tissue types represent more or less sharp edges. Gibbs phenomena are most pronounced for the quasi realistic simulation model.

3.7.2 Insufficient edge-preservation

Primarily, the ability to edge preservation correlates with the density of the knot partition. However, an arbitrarily intense agglomeration of the basis functions seems to be useless as evident from Fig. 3.12. Let the gray-colored square represent a 2d area of closely packed details, exposing a sharp edge to the rather smooth background displayed in white. The non-uniform 1d distributions of knots ξ_i and κ_j aim at modelling the gray zone with enough 2d knots. As disturbing side effect, more knots than intended are also placed in the background (see points (ξ_6, κ_j) and $(\xi_i, \kappa_7), i = 1, \ldots, 6, j = 1, \ldots, 7$). Hence, neither the multidimensional nor the sequential smoothing variant of the SVCM approach allow

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

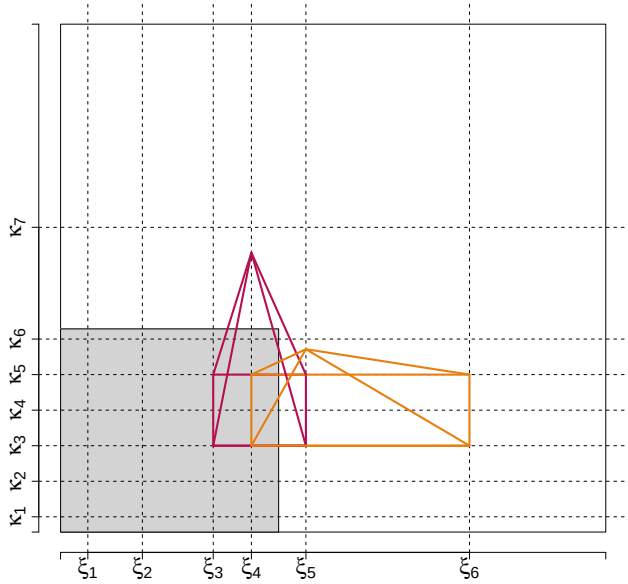


Figure 3.12: This scheme clarifies two aspects: a) non-uniformly distributed 1d knots can lead to higher 2d knot density where needed (gray area) but at the expense of additional knots in the presumably flat background (white); b) two adjacent basis functions can only model the intermediate egde when local penalization is relaxed.

for arbitrary knot setting. The use of geometrically more flexible basis functions such as radial basis functions, non-uniformly distributed rational B-splines (NURBS; in the context of DTI see Weickert and Hagen (2006, Chap. 18)) or wavelets might be more suitable to improve local adaptiveness around biological "steps".

As promising alternative, local adaptiveness can be increased by introducing data-driven penalty weights. Consider, for instance, the two schematic linear B-splines that lie in direct neighborhood across the edge (Fig. 3.12). A large difference of their amplitudes is mandatory to work out the edge. Yet, the application of first or even second order difference penalties runs contrary to this assumption. Therefore, the corresponding penalty must be downweighted. Suitable weights can be generated from approximations to the first or second derivative using e. g. Sobel or Laplace filter. The voxelwise local gradients within the support region of each basis function can be comprised into one single measure, which then serves directly or reciprocally as weight for the basis function. Note, however, that this procedure results in a two-stage estimation process. A related concept was pub-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

lished by Sinha and Schunk (1992) who used moving least median of squares to obtain an intermediate robust surface, subsequently derived weights according to discontinuity preservation and finally fitted the data using weighted bicubic splines.

3.7.3 Star-shaped artifacts

Feature preservation is paramount in DTI regularization. Hence, the use of linear B-splines with their smaller support regions aims at local adaptiveness. However, linear B-splines are apparently far from being isotropic as the support in 3d becomes tetrahedra-like. This causes star-shaped artifacts, particularly annoying when it comes to interpolation. Bright spots, i. e. blips, can be recognized from the whole brain slice image as well as from the closeup (Fig. 3.13).

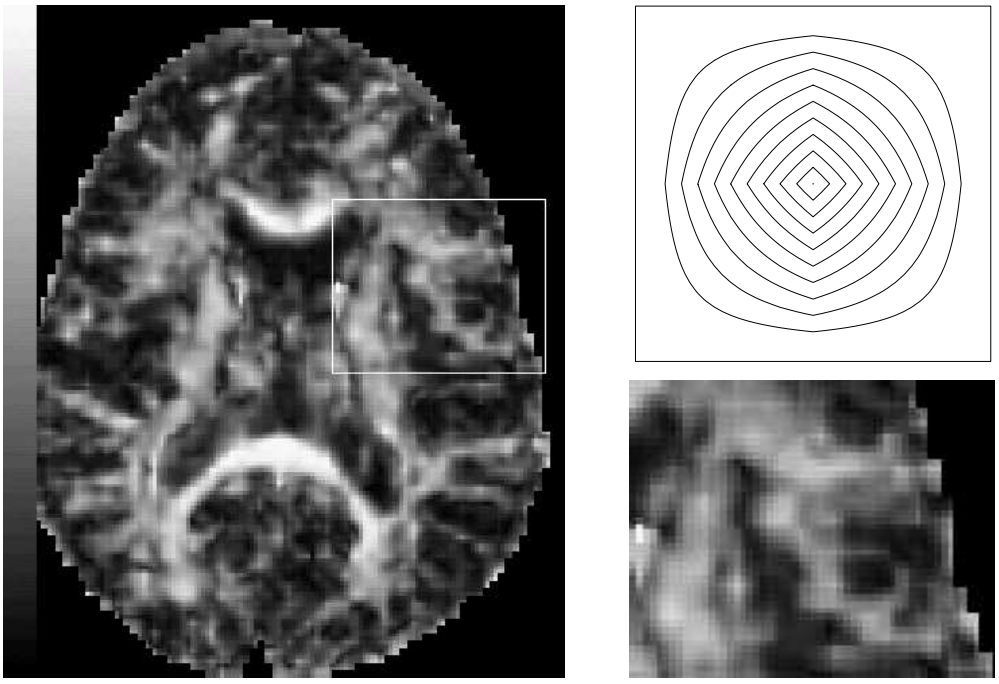


Figure 3.13: FA map of an exemplary brain slice (left) and closeup of the superimposed squared region (right). The FA values are derived from a 3d tensor field, estimated and interpolated to double size using SVCM with linear B-splines and first order difference penalties. The bright star-shaped artifacts can be attributed to the shape of the basis functions, depicted top right for 2d.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3.8 Implementational issues

As the straightforward approach to SVCM (Eq. (3.9)) runs into difficulties, we can alleviate computation by taking advantage of the repeated measures structure, i. e. $\mathbf{X}$ does not vary across voxels. Since the covariates are on a grid, the expectation of $\mathbf{Y}$ can be expressed as a tensor product involving $\mathbf{X}$. Hence, we could apply the fast, compact smoothing algorithm of Eilers et al. (2006). Such an approach avoids the computation of large Kronecker products of B-spline bases, and with a trick, this algorithm turns the 3d VCM into smoothing with 4d tensor products. The limiting bottleneck remains the size of the equation system to be solved, which consists of several tens of thousands of normal equations in a realistic scenario. The relevant left hand side, namely $\mathbf{U}'\mathbf{U} + \mathbf{P}$, would occupy more than 18 Gb for a realistic number of parameters, i. e. $6 \cdot 32 \cdot 32 \cdot 8 = 49,152$.

We achieve a crucial storage gain by exploiting the sparsity of the involved matrices, making the brain SVCM model tractable again. For illustration purpose, Figure 3.14 includes a spy diagram of row, column, layer, and complete penalty matrices with non-zero entries colored black.

Beyond the conspicuously large amount of zeros, the banded structure strikes at first glance. Since the matrix $\mathbf{B}$ of tensor product basis functions is also sparse and exhibits a band pattern, $\mathbf{U} = \mathbf{B} \otimes \mathbf{X}$ and $\mathbf{U}'\mathbf{U}$ inherit these two properties being unaffected by the addition of the complete penalty. Therefore, the use of suitably sparse matrix libraries is indispensable. For example, the aforementioned scenario then becomes feasible again with a required amount of 53 Mb working memory for storing the left hand side of the corresponding equation system. Note that combination with array regression as decsribed

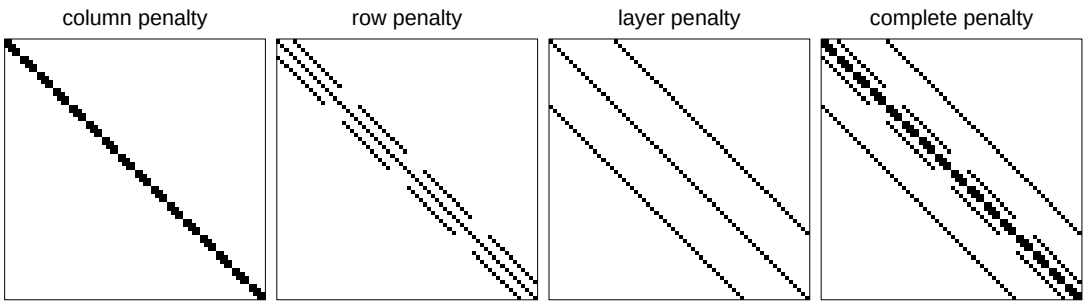


Figure 3.14: Penalty matrices corresponding to $3 \times 3 \times 3$ knots, quadratic B-splines, first order of difference penalties.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

in Eilers et al. (2006) and Currie et al. (2006) would require a compressed format for sparse 4d arrays which is, to the best of our knowledge, not available with current software packages.

Concerning the optimization of the smoothing parameter, a greedy grid search with look-up table was performed in the dimension specific case while a combination of golden section search and successive parabolic interpolation (as available in R; R Development Core Team (2006)) was applied in the one-dimensional case. The greedy grid search finds the minimum GCV correctly on a refined $10 \times 10 \times 10$ log-scaled grid with equidistant exponents, λ_1 and λ_2 varying from 10^{-3} to $10^{0.5}$, and λ_3 within the range of 10^{-8} to 10^{-3}. It has to be pointed out that fixing the tuning parameter to the average over the optimal parameter values corresponding to data sets from a small sample of representative healthy subjects, is expected to satisfy the application of our proposed SVCM in daily use. Hence, the time-expensive optimization is dispensable in practice.

All routines are implemented in R (R Development Core Team, 2006) and available as package `svcm` from the archives
(URL http://cran.r-project.org/src/contrib/PACKAGES.html). To this end, the package exploits the band structure of the hat matrix for computing the GCV according to Hutchinson (1985). As an added benefit, `svcm` runs without installation of additional software. However, analyses reported in this paper were conducted on the basis of sparse matrix formats incorporating the library Taucs (Toledo et al., 2003), running about three times faster.

3.9 The 'svcm'-package

As pointed out in the Section 3.8, routines have been made available for conducting both the tensor product and the sequential SVCM approach. The software is realized within the open-source platform that is probably the most commonly known to statisticians, above all to those active in research (R Development Core Team, 2006). Since the package provides 2d as well as 3d example data and appropriate example calls, its publication follows the concept of reproducible research.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

3.9.1 Getting started

Having launched R, the package `svcm` can easily be installed from CRAN using the command:

```
R> install.packages("svcm", dependencies=TRUE)
```

The optional argument serves to guarantee the automatic installation of the required packages `Matrix` (Bates and Maechler, 2006) and `splines` (R Development Core Team, 2006). In a next step, the package needs to be loaded into the current R session:

```
R> library(svcm)
```

To view the documentation on the provided data (`brain2d, brain3d`) and functions (`cleversearch, svcm, resolution`) use the help command, for example `?brain2d`.

3.9.2 Data format

In detail, the data sets consist each of six diffusion weighted images (DWI) of the human brain. While the 2d images show a complete representative axial slice of 90×75 voxels, the 3d images contain a small contiguous region of the original whole brain volume. The extract of $15 \times 30 \times 6$ voxels depicts the posterior part of the lateral ventricles and the corpus callosum. Hence, areas with both low and high signal intensities are included. Either data sets are preprocessed according to model (3.4) in Section 3.2 and formatted as 3d and 4d arrays, respectively. The last dimension encodes for the direction of the six applied diffusion weighting gradients. More details on the specific acquisition settings and the data origin are attached to the help files.

For data import, e. g. of the 3d data set, use the `data` command

```
R> data(brain3d)
```

The example section also supplies the user with an appropriate visualization call leading to Fig. 3.15 in the 3d case and to Fig. 3.17 (a) in the 2d case.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

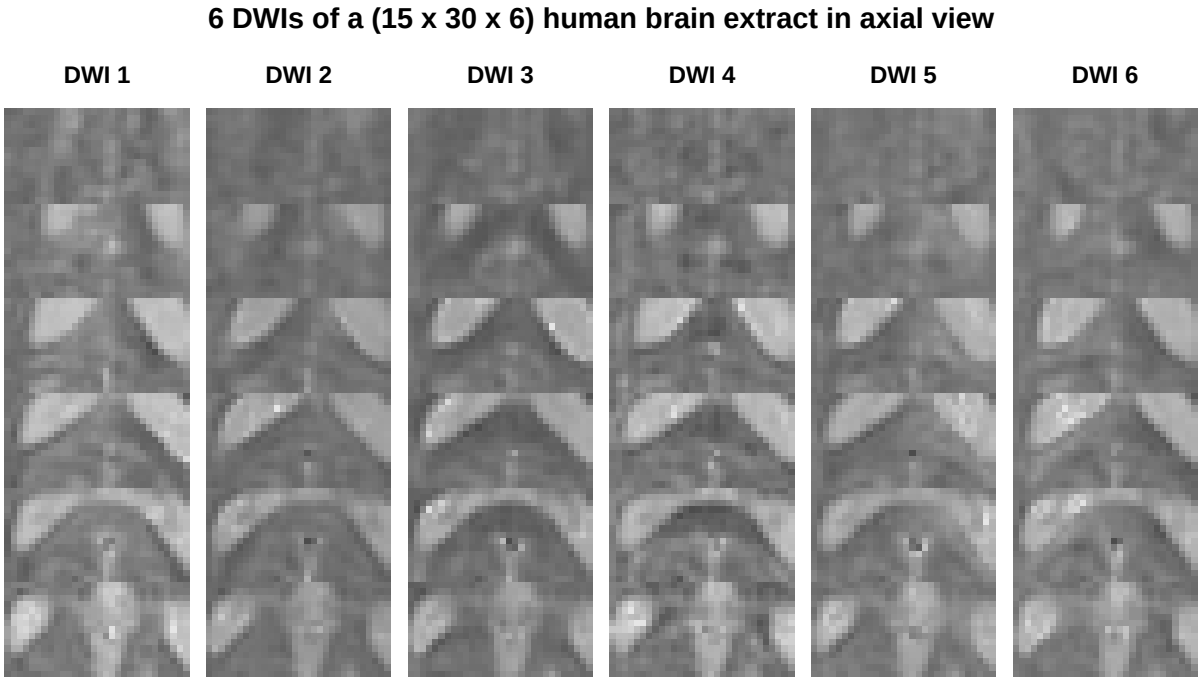


Figure 3.15: Generated using example call in `?brain3d`.

3.9.3 SVCM analysis

The core routine is described in `?svcm`. Specification of the 3d or 4d array `Y` of observational data and the regressor matrix `X` is mandatory. Also, an initial value needs to be assigned to the tuning parameter. Further optional arguments concern, on the one hand, the voxel size, the number of knots, the B-spline degree, the penalty order, and the SVCM variant to use. On the other hand, optimization specific parameters might be passed such as whether to perform optimization at all and if so, whether this should happen on a multidimensional grid or on a global scale.

An important prerequisite is the model assumption of space-invariant regressors. Moreover, data must be on a regular grid without missings. The observations in `Y` are assigned to the center of the respective grid unit. As a consequence, the basis functions are evaluated at these coordinates.

The function can also act as pure smoother for `Y` if the user sets `X = matrix(1, 1, 1)` and ascertains that the last dimension of `Y` matches `dim(X)[1]`, i. e. if the user does not drop pseudodimensions of `Y`. Note that the last dimension of `Y` must always correspond to the number of rows of `X`. The columns of `X` determine how many space-varying coefficients

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

are to be estimated.

Following the given illustrative material, import the 3d brain data and specify the corresponding regressor matrix as in Section 3.2:

```
R> X <- matrix(c(0.5, 0.5,   0,   0, 0.5, 0.5,
                   0,   0, 0.5, 0.5, 0.5, 0.5,
                 0.5, 0.5, 0.5, 0.5,   0,   0,
                   0,   0,   0,   0,   1,  -1,
                   1,  -1,   0,   0,   0,   0,
                   0,   0,   1,  -1,   0,   0), nrow = 6)
```

Then estimate the SVCM using 3d tensor products (`type="TP"`) of linear B-splines (`deg=c(1, 1, 1)`) as well as a fixed (`search=FALSE`) and global tuning parameter (`lambda.init=1`):

```
R> M3d <- svcm(brain3d, X, knots=c(5, 10, 5), deg=c(1, 1, 1), type="TP",
               vsize=c(1.875, 1.875, 4.0), search=FALSE, lambda.init=1)
```

The procedure returns a list that contains, among others, the respective arrays of fitted values, effects and coefficients as exemplarily displayed in Fig. 3.16.

3.9.4 A note on optimization

If optimization is desired, it is governed by the GCV criterion (Hastie and Tibshirani, 1990):

$$\mathrm{GCV}(\boldsymbol{\lambda}) \quad = \quad n\frac{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}{n-\mathrm{tr}(\mathbf{H})}.$$

The computation involves the diagonal of the hat matrix $\mathbf{H}, \mathbf{H} = \mathbf{U}(\mathbf{U}'\mathbf{U}+\mathbf{P})^{-1}\mathbf{U}'$, $\mathbf{U} = \mathbf{B} \otimes \mathbf{X}$ (compare Section 3.3.2). Yet in some cases, the inversion, $\mathbf{H}$ is based on, might become almost unfeasible with current computers since the matrix $\mathbf{S} = \mathbf{U}'\mathbf{U} + \mathbf{P}$ can quickly grow in size depending on the number of knots, the spline degree and the order of difference penalties. In the sequential variant of SVCM this does not represent a problem as obvious from Eq. (3.17) where $\mathbf{H}$ is restated as Kronecker product of dimension specific hat matrices. Hence, the diagonal of $\mathbf{H}$ equals the Kronecker product of the

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

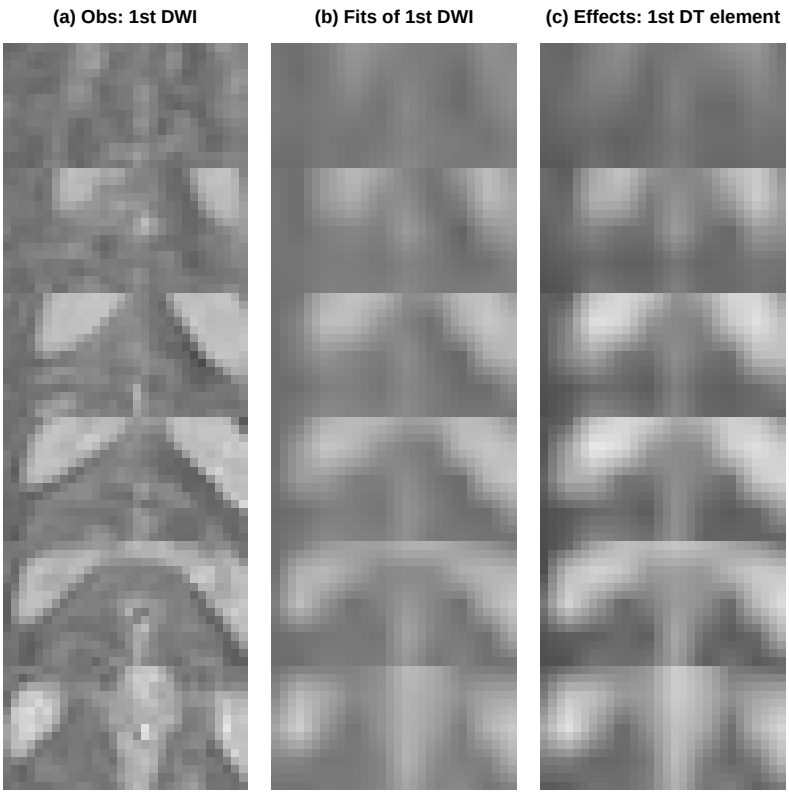


Figure 3.16: Six axial slices of the 1st DWI sample (a) and its fit (b); (c) corresponds to the first component of the diffusion tensor.

respective diagonal matrices and can efficiently be computed using sparse matrix formats. Furthermore, the trace of Kronecker products is the product of the traces.

Yet in the tensor product approach, suitable workarounds are badly needed to overcome physical limitations (RAM, available access memory). The one that is also included in the package bases on a recursion formula for band matrices Hutchinson (1985). Thomas Kneib provided the basic C code and Michael Kobl built the ready-for-use shared object. The crucial quantity is the size of the envelope, i. e. the vector of the off-diagonal elements starting from the first non-zero row entries (George and Liu, 1981). As obvious from Fig. 3.14, multiple Kronecker products in $\mathbf{U}$ and $\mathbf{P}$ can cause a wide envelope. Hence, preparatory fill-reducing ordering of $\mathbf{S} = \mathbf{U}'\mathbf{U} + \mathbf{P}$ according to e. g. Cuthill-McKee (George and Liu, 1981) could lead to a considerable speed-up.
A powerful alternative exploits once more the extreme sparsity of design matrix. This approach is however not distributed with the R-package due to its dependence on Taucs, a library of sparse linear solvers. The idea is to reformulate the inversion problem as

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

system of linear equations. The derivation starts from the original task of determining the trace of the hat matrix:

$$\begin{aligned}\mathrm{tr}(\mathbf{H}) &= \mathrm{tr}(\mathbf{U}\mathbf{S}^{-1}\mathbf{U}') \\ &= \mathrm{tr}(\underbrace{\mathbf{S}^{-1}\mathbf{U}'\mathbf{U}}_{=\mathbf{\Phi}}) \\ &= \sum_{i=1}^{pKLM} \mathbf{\Phi}(i,i),\end{aligned}$$

with $(rn_1n_2n_3 \times pKLM)$-matrix $\mathbf{U}$ and $(pKLM \times pKLM)$-matrix $\mathbf{S}$, $pKLM < rn_1n_2n_3$. The unknown elements in $\mathbf{\Phi}$ can be assessed by solving the sparse linear equation system

$$\mathbf{S}\mathbf{\Phi} = \mathbf{U}'\mathbf{U}$$

of dimension $(pKLM)^2$. Depending on the knot density, this can still be close to the dimension $rn_1n_2n_3$ of the original problem. To reduce the required memory, optimization concentrates on the diagonal of $\mathbf{\Phi}$. Hence, loop over $i, i = 1, \ldots, r$, and call Taucs' linear solver for sparse systems (Toledo et al., 2003) on each column of the parameter matrix $\mathbf{\Phi}$:

$$\mathbf{S}\mathbf{\Phi}(\cdot, i) = (\mathbf{U}'\mathbf{U})(\cdot, i),$$

thereby storing the diagonal elements $\mathbf{\Phi}(i,i)$ only. Nicely, the underlying factorization of $\mathbf{S}$ needs to be performed only once, allowing for additional speed-up of the inversion.

In case of a global tuning parameter, optimization can be performed using the R internal routine `optimize`, requiring an interval to be passed as parameter space. Attention should be paid on the restriction $\lambda \geq 0$. Whenever the tuning parameter is dimension specific, the simplex method "Nelder-Mead" or quasi-Newton "L-BFGS-B" with positivity constraint for the smoothing parameter are conceivable specifications for the internal function `optim`. Yet, a greedy or optionally full grid search can alternatively be employed, offering advantageous control over the parameter space. Since this latter routine could be of general interest, it constitutes an autonom function named `cleversearch`. The arguments encompass the objective function and several grid settings such as the dimension, the lower and upper bounds, and whether these should be on a logarithmic scale. For demonstration purpose, call the sequential SVCM type with the 2d example data and a 2d grid search for the smoothing parameter by

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

```
R> M <- svcm(brain2d, X, knots=c(60, 50), deg=c(1, 1), search=TRUE,
             type="SEQ", lambda.init=rep(0.1, 2), vsize=c(1.875, 1.875),
             method="grid", lower=rep(-5, 2), upper=rep(0, 2), ngrid=10)
```

The results are shown in Fig. 3.17 (b) and (c).

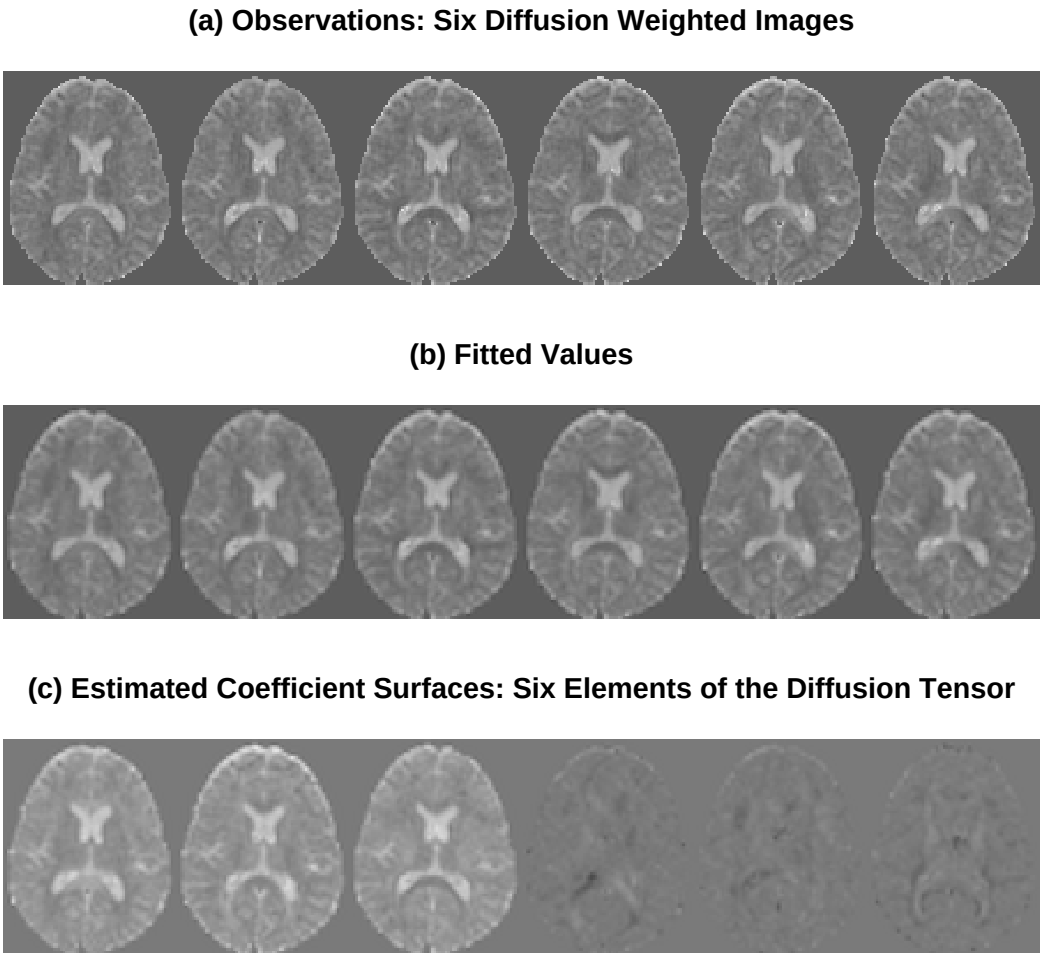


Figure 3.17: (a) results from example call given in `?brain2d`*, (b) and (c) are produced as demonstrated in* `?svcm`*.*

3.9.5 Resolution increment

The last function of the package (`resolution`) serves post-hoc adjustment of the resolution of SVCM derived predictors and effects. Naturally, the mandatory arguments encompass the output list from an `svcm` function call, the corresponding regressor matrix `X` and a rescaling factor denoted by `fac`.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

The intuition suggests that resolution doubling would return values at the former locations plus intermediate points. In order to alter the original resolution, the basis functions are evaluated for a correspondingly larger (or smaller) set of points. This raises the question of coordinate assignment. An adequate assumption is that the voxel center is most representative for the whole voxel. Then, for instance, resolution doubling would intuitively be expected to result in values at the former locations plus intermediate points. It is readily understood that equidistance is no longer given after such a transformation. However, for equidistance being strongly desired, the `fac`-times resolution of one dimension with n voxels of size `vsize` is based on coordinates

$$(i - 0.5) \cdot \texttt{vsize}/\texttt{fac}, i = 1, \ldots, n \cdot \texttt{fac}.$$

This means, estimates at original grid points do not enter the resized arrays as visualized in Fig. 3.18, at least for double or triple resolution increment. The formula is applied into x-, y-, and z-direction and results in a refined equidistant 2d and 3d grid, respectively.

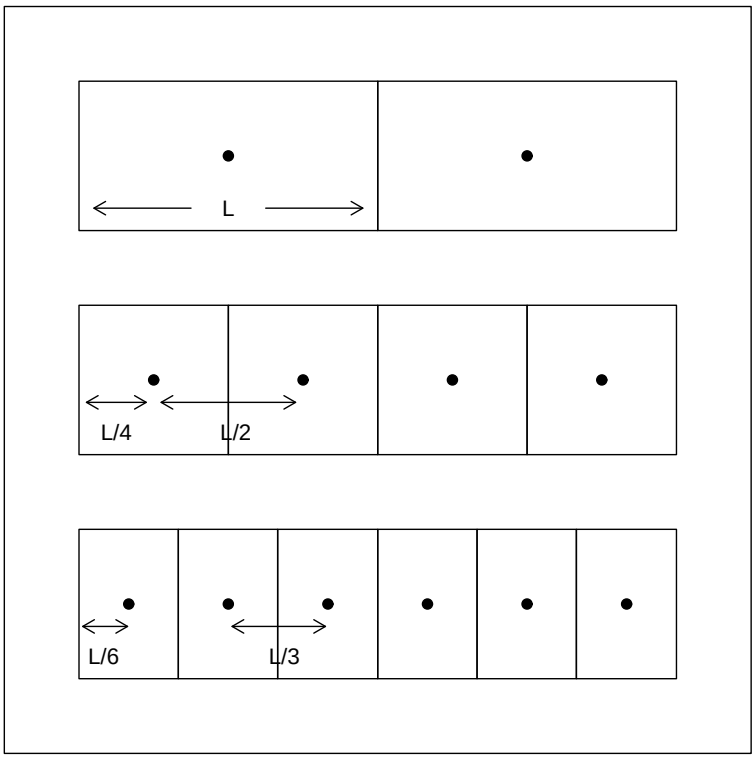


Figure 3.18: Original, double and triple resolution scheme (top to bottom). Every component of `vsize` *determines a dimension specific voxel length L.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Figure 3.19 can be produced according to the example call:

```
R> M2 <- resolution(X, M, fac=c(2, 2))
```

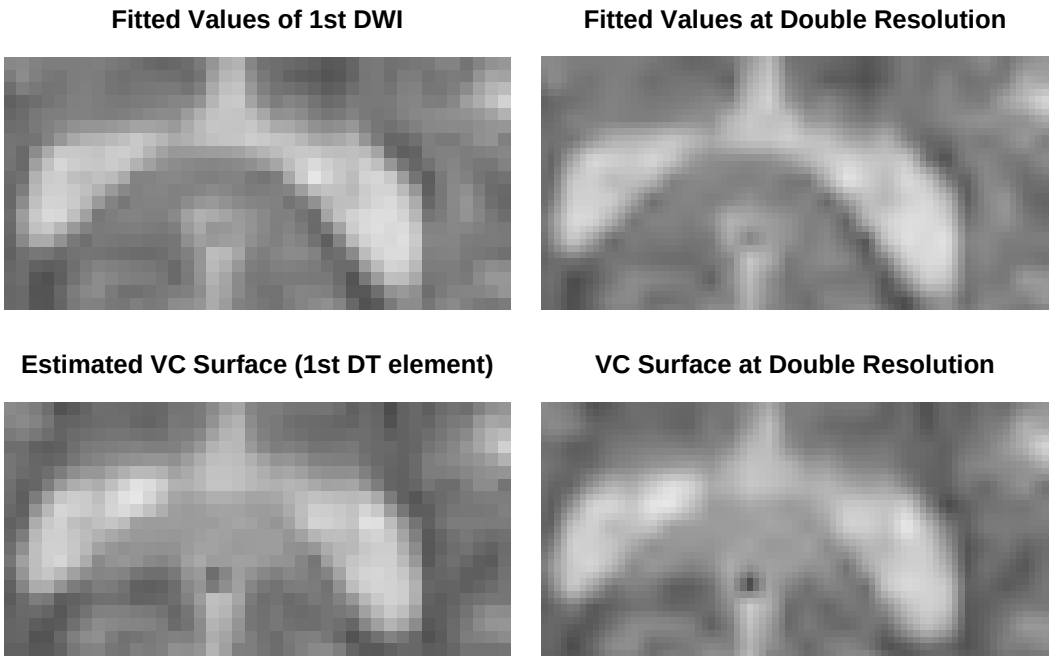


Figure 3.19: Exemplary predicted region (top row) as well as one space-varying coefficient surface over the same region at original and double resolution.

3.9.6 Possible extensions

In principle, confidence intervals can be assessed for the coefficients, the effects and the fitted values. Given the SVCM equation (3.5) in Section 3.3 with the space-varying coefficients $\boldsymbol{\beta}_j$ projected onto tensor product B-splines as in Eq. (3.6), the penalized coefficients have the solution $\hat{\boldsymbol{\gamma}} = (\mathbf{U}'\mathbf{U} + \mathbf{P})^{-1}\mathbf{U}'\mathbf{y}$ (see Eq. (3.9)). Thus, the covariance matrix is given by

$$\mathrm{Cov}(\hat{\boldsymbol{\gamma}}) \quad = \quad \hat{\sigma}^2(\mathbf{U}'\mathbf{U} + \mathbf{P})^{-1}\mathbf{U}'\mathbf{U}(\mathbf{U}'\mathbf{U} + \mathbf{P})^{-1}.$$

A simplifying and more conservative covariance expression for $\hat{\boldsymbol{\gamma}}$ is stated in e. g. Hastie and Tibshirani (1990, Chap. 3.8). The reasoning follows from a Bayesian point of view on the underlying model:

$$\mathbf{y} \quad = \quad \mathbf{U}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I})$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

with a prior $\boldsymbol{\gamma} \sim N(0, \sigma^2 \mathbf{P}^-)$ and $\mathbf{P} = \lambda_1 \mathbf{P}_1' \mathbf{P}_1 + \lambda_2 \mathbf{P}_2' \mathbf{P}_2 (+\lambda_3 \mathbf{P}_3' \mathbf{P}_3)$ derived from the 2d or 3d penalty term (compare Eq. (3.8)). Then the a posteriori distribution $\boldsymbol{\gamma}|\mathbf{y}$ is multivariate normal with mean $(\mathbf{U}'\mathbf{U} + \mathbf{P})^{-1}\mathbf{U}'\mathbf{y}$ and covariance

$$\mathrm{Cov}(\boldsymbol{\gamma}|\mathbf{y}) \;=\; \sigma^2 (\mathbf{U}'\mathbf{U} + \mathbf{P})^{-1}. \tag{3.21}$$

An estimate of σ^2 is given by

$$\hat{\sigma}^2 \;=\; \frac{\|\mathbf{y} - \mathbf{U}\hat{\boldsymbol{\gamma}}\|^2}{n - \mathrm{tr}(\mathbf{H})},$$

where $\mathbf{H} = \mathbf{U}(\mathbf{U}'\mathbf{U} + \lambda \mathbf{P})^{-1}\mathbf{U}'$ denotes the hat matrix as before. Using Eq. (3.21), covariance matrices of effects and fitted values can be computed according to:

$$\begin{aligned} \mathrm{Cov}(\mathbf{B}\hat{\boldsymbol{\gamma}}) &= \hat{\sigma}^2 \mathbf{B}\, \mathrm{Cov}(\hat{\boldsymbol{\gamma}}) \mathbf{B}', \\ \mathrm{Cov}(\mathbf{U}\hat{\boldsymbol{\gamma}}) &= \hat{\sigma}^2 \mathbf{U}\, \mathrm{Cov}(\hat{\boldsymbol{\gamma}}) \mathbf{U}' = \hat{\sigma}^2 \mathbf{H}. \end{aligned}$$

In fact, only the diagonal elements are of interest in order to compute the respective $(1-\alpha)$-confidence intervals:

$$\begin{aligned} \mathrm{CI}(\boldsymbol{\beta}) &= \left\{ \hat{\boldsymbol{\beta}} \pm u_{1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2\, \mathrm{diag}(\mathbf{B}\, \mathrm{Cov}(\hat{\boldsymbol{\gamma}}) \mathbf{B}')} \right\} \\ \mathrm{CI}(\boldsymbol{\mu} = \mathrm{E}(\mathbf{y})) &= \left\{ \hat{\boldsymbol{\mu}} \pm u_{1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2\, \mathrm{diag}(\mathbf{H})} \right\}, \end{aligned}$$

where $u_{1-\frac{\alpha}{2}}$ denotes the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution.

Acknowledgements

We gratefully acknowledge Thomas Kneib for valuable comments, and Philipp Sämann for advice on neuroscientifical applications. The authors thank also the anonymous reviewers for their insightful comments and suggestions as well as the Max Planck Institute of Psychiatry (Munich, Germany) for providing us with human DTI data. This study was kindly supported by the collaborative research center SFB386 *Statistical Analysis of Discrete Structures* as part of the German Science Foundation (D. F. G.). The research of Brian Marx was partially supported by National Science Foundation grant: IIS-0326387 and AFOSR grants: F49620-03-1-0238, F49620-03-1-0239, and F49620-03-1-0241, FA9550-05-1-0454.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4 Wavelets for DTI

In this chapter, wavelet basis functions are investigated for their suitability for processing and analysing DTI data. First, wavelet theory is introduced and explained by means of 1d and 2d examples (Sections 4.1.1 – 4.1.3). General thresholding techniques, which serve as regularization concepts for wavelet based models, are presented in Section 4.1.4. Regularization of DTI data can be performed at two stages, either immediately after acquisition (Wirestam et al., 2006) or after tensor estimation. The latter stage of denoising is outlined in Section 4.6 together with the incorporation of the positive definiteness constraint using log-Cholesky parametrization. In Section 4.3, the procedure is examined in a simulation study and compared to standard processing and the space-varying coefficient model (SVCM) based on B-splines (Eq. (3.5) – (3.7)). In addition, a real data example is presented and discussed. Finally, an approach is proposed how a space-varying coefficient model could fairly be adapted to wavelet basis functions. The theoretical parts are based on books of Gençay et al. (2002, Chap. 1, 4-6), Härdle et al. (1998), Ogden (1997) and Jansen (2001) if not stated otherwise.

4.1 Introduction to wavelets

Before defining wavelets and proceeding with their applicability to DTI data, I would like to reason why the use of wavelets seems to be promising. Without loss of generality, space will be the location index below, comprising time as special 1d space.

Looking at Fig. 4.1, the partition of the space-frequency plane is displayed for different transformations. B-splines, as used in the previous chapter, are only localized in the space domain (Fig. 4.1 (a)). For example, zero-degree B-splines are piecewise constants

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

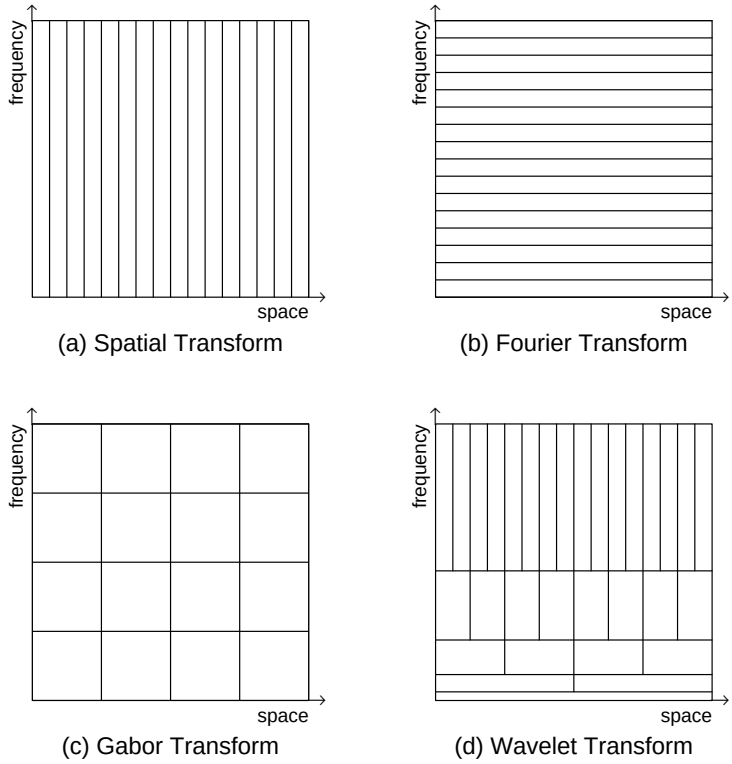


Figure 4.1: Partition of the space-frequency plane by different transforms.

covering one single interval. Hence, a 1d sequence of DTI signal intensities can fully be represented by as many zero-degree B-splines as the sequence contains observations. In contrast, a Fourier analysis using cosine and sine functions offers full resolution in the frequency domain only (Fig. 4.1 (b)). Additional coarse localization in space can be gained by windowing the original series and applying the Fourier analysis across each window (Fig. 4.1 (c)). This modification is also known as short-time Fourier transform or Gabor transform after its originator. However, spatial resolution is increased at the expense of frequency resolution and spatial localization is still lacking within the fixed windows. This phenomenon is due to Heisenbergs uncertainty principle, stating that a particle cannot exactly be described by its position and momentum at the same time. Wavelet transforms, by contrast, intelligently adapt to the space-frequency plane in that a narrow space-window is provided to examine high-frequency content, but a wide space-window is allowed when investigating low-frequency components (Fig. 4.1 (d)). Due to their resulting localization in time/space *and* in frequency the use of wavelets promises local adaptiveness. In other words, the wavelet transform can be expected to zoom in higher frequencies when distinctive spikes or sharp edges are met.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

The switch to wavelets is furthermore motivated by recent reports about successful wavelet based filtering of MR images with best contrast rendering between different tissue types and best spatial adaptivity. A selection of articles on the application of wavelets to radiological imaging can be found in the special issue of the IEEE Journal on Transactions in Medical Medicine (March 2003). It was however in the early nineties when Weaver et al. (1991) initially proposed the removal of random noise from MR images using selective wavelet reconstruction which Donoho and Johnstone (1994) developed to a full statistical technique. There is also a series of papers using wavelets in order to denoise fMRI data as well as to enhance the power of statistical analyses, see e. g. Van de Ville et al. (2006) and the references therein. Substantial contributions to the processing of MR modalities also stem from the Biomedical Imaging Group (École Polytechnique Fédérale, Lausanne). See the steadily updated homepage[1] for the most recent publications.

4.1.1 Univariate wavelet basis

With the aim in view to design space-varying coefficient surfaces using 3d wavelet basis functions, let me launch from the classical, more intuitive one-dimensional case. To give a first impression, Fig. 4.2 shows different basis functions of the Daubechies extremal phase wavelet class.

There is always a pair of so-called *father* and *mother* wavelet functions, $\phi(x)$ and $\psi(x)$, respectively. The father wavelet is also called *scaling function*, while the mother wavelet is simply entitled *the* wavelet due to its prototype functionality for an entire set of wavelets. As most essential condition and kind of wedding ring, father and mother wavelet are orthogonal to each other. For the Haar family, this can immediately be seen from the graphic (top row) and, mathematically, by the corresponding definitions of the scaling function

$$\phi(x) = \begin{cases} 1, & x \in [0,1) \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

[1] http://bigwww.epfl.ch/

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

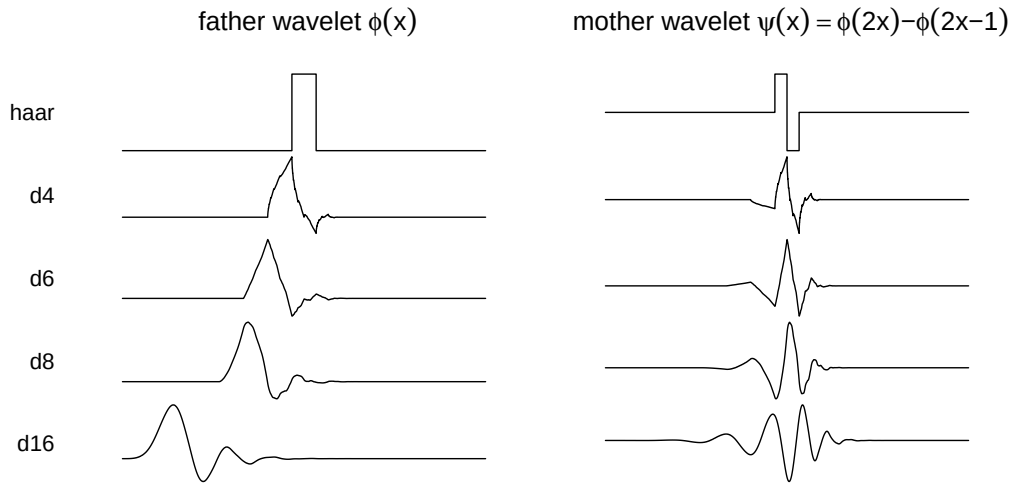


Figure 4.2: Principal representatives of the Daubechies extremal phase wavelet class.

and the wavelet function

$$\psi(x) \quad = \quad \phi(2x) - \phi(2x-1) = \begin{cases} 1, & x \in [0, 0.5) \\ -1, & x \in [0.5, 1) \\ 0, & \text{otherwise}\,. \end{cases} \tag{4.2}$$

As furthermore obvious from Fig. 4.2, the representatives differ in (compact) support length and the degree of smoothness. Apart from the Haar family, asymmetry holds for the given wavelets and, in general, for most of the known families. Among the tremendous number, the main families are Daubechies extremal phase wavelets, Daubechies linear phase wavelets (also known as least asymmetric wavelets or simply symlets) and coiflets. Some other classes involve B-splines such as the Battle-Lemarié and Chui-Wang cardinal spline families (both explained in Ogden (1997, pp. 23)), and the orthogonal fractional B-spline wavelets (Blu and Unser, 2003). Interestingly, the Haar wavelet cannot only be regarded as member of the Daubechies extremal phase but also of the Battle-Lemarié family. For an excellent description of how to construct wavelet bases see e. g. Härdle et al. (1998, Chap. 5-6).

Given its simplicity, the Haar wavelet will serve as the toy wavelet throughout this chapter. Most other wavelet classes do not even possess an explicit space-domain formula and are only defined implicitly.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

There are two formal conditions for a function $\psi(x), x \in \mathbb{R}$, to be a wavelet. The *admissibility condition*

$$\int_{\mathbb{R}} \psi(x)dx = 0$$

ensures integration to zero and existence of the inverse. It is equivalent to a zero-valued zeroth moment. Further vanishing moments are reflected by the family attribute number, e. g. Daubechies-4 wavelets expose four vanishing moments. Seldom the father wavelet possesses vanishing moments, too, as it is the case for coiflets (Härdle et al., 1998). Note that some authors let the attribute encode for the length of the mother wavelet which is twice the number of vanishing moments. As evident from Fig. 4.2, the smoothness property is more pronounced for wavelets with larger support, as there is more time/space to die out to zero.

The second condition, namely the *unit energy condition*,

$$\int_{\mathbb{R}} \psi^2(x)dx = 1$$

constrains $\psi(x)$ to have normalized strength. The term *energy* refers to the "size" of a signal. Whenever a signal is imagined as a function of varying amplitude through time, the area under the curve seems to reason a good measurement of the signal strength. Since the possibly negative parts do not have less strength than a positive signal of the same size, energy is defined as the area under the squared signal. Hence the unit energy condition forces $\psi(x)$, given its admissibilty, to take on nonzero values for some point x. As a consequence any wavelet function will oscillate or wave around the abscissa. The suffix 'let' comes in from discrete dilation and translation of the mother's shape, eventually resulting in baby wavelets

$$\psi_{j,k}(x) = \frac{1}{\sqrt{2^{-j}}}\psi\left(\frac{x - k2^{-j}}{2^{-j}}\right) = 2^{j/2}\psi(2^j x - k), \quad j \in \mathbb{Z}_0^+, k \in \mathbb{Z} \tag{4.3}$$

with support $[k2^{-j}, (k+1)2^{-j})$.

Wavelets are thus functions in two parameters. Parameter k accounts for dyadic translational displacement in the time domain whereas index j constitutes a dyadic change in frequency. The latter corresponds to compression or stretching of the wavelet on the time axis as illustrated by selected representatives in Figure 4.3. The amount of "localization" is therefore controlled by the dilation parameter j.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

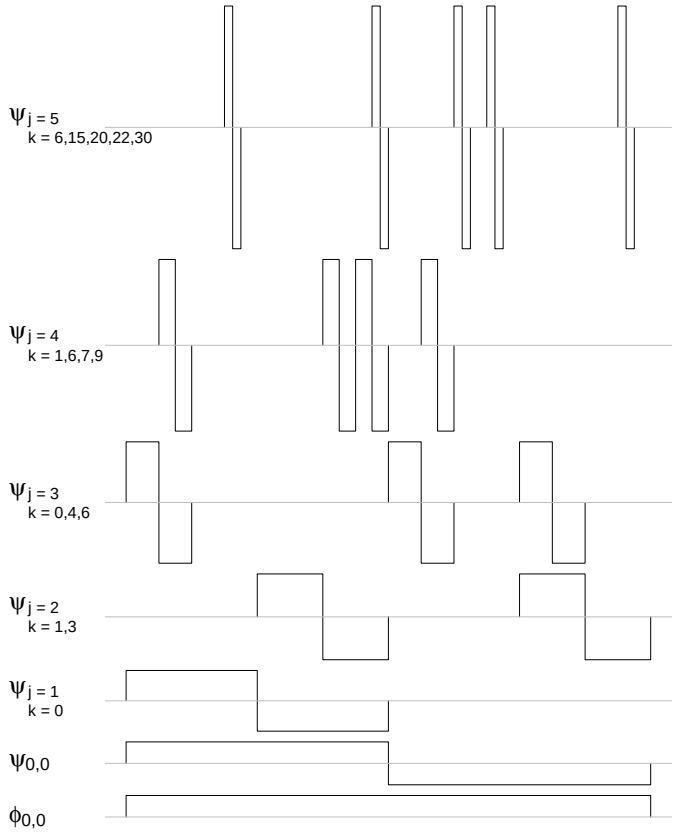


Figure 4.3: Selected members of the Haar family.

Application of these simple operations, namely dilation and translation, to the mother wavelet and multiplication with a normalization constant $2^{j/2}$ naturally leads to an orthonormal system

$$\left\{\{\phi_{0,k}\}, \{\psi_{j,k}\}, k \in \mathbb{Z}, j \in \mathbb{Z}_0^+\right\}$$

in $L^2(\mathbb{R})$, the space of square integrable functions[2] (Härdle et al., 1998). The system of wavelet functions allows to approximate any function $f \in L^2(\mathbb{R})$ aribitrarily well. For a pre-assigned precision, the system turns into an orthonormal basis for a subspace of $L^2(\mathbb{R})$. The shifted versions $\phi_{0,k}(x) = \phi(x-k)$ guarantee the coverage of $\mathbb{R}$. An outline of the theoretical underpinnings are postponed to the next section.

Orthogonality of the basis functions is achieved via critical sampling of the space-frequency plane at the nodes in Fig. 4.1 (d). This way the continuous and redundant partitioning becomes discrete and orthogonal with a minimum number of discrete points in space $k2^{-j}$

[2] A function f is square integrable or of finite energy if $\int_{\mathbb{R}} f^2(x)dx < \infty$.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

and discrete frequencies 2^{-j}. Wavelets obtained by Eq. (4.3) are therefore named discrete wavelets though being (piecewise) continuous functions.

4.1.2 Wavelet decomposition and reconstruction

In analogy to the Fourier or the zero-degree B-spline transform, any function $f \in L^2(\mathbb{R})$ can be approximated by a linear combination of wavelets

$$f(x) = \sum_{k=-\infty}^{\infty} \alpha_{0k}\phi_{0k}(x) + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} \beta_{jk}\psi_{jk}(x),$$

with cofficients α_{0k} and β_{jk} (Härdle et al., 1998). This discrete decomposition directly results from the basis of discrete wavelets claimed in Section 4.1.1 which will roughly be reasoned in the following. In practice, only a discretized version of the function of interest is observed and this, in general, under some additional noise. Assume the moreover finite number of measurements to be equally spaced on the support region of dyadic length, say $[0, n = 2^J)$. This constitutes a square-summable[3] function f^J in the space V_J, defined as

$$V_J = \{f \in l^2(\mathbb{Z}) : f \text{ is piecewise constant on } [k2^{-J}, (k+1)2^{-J}), k \in \mathbb{Z}\}.$$

Here, $l^2(\mathbb{Z})$ denotes the space of square-summable functions with integer domain. To gain a better understanding of how the concept of the discrete wavelet transform is derived subsequently, consider $J = 3$ and Figure 4.4.
An approximation to f^J can be generated by a member f^{J-1} of the subspace

$$V_{J-1} = \{f \in l^2(\mathbb{Z}) : f \text{ is piecewise constant on } [k2^{-J+1}, (k+1)2^{-J+1}), k \in \mathbb{Z}\}.$$

As the *level* $j, j = 0, \ldots, J$, decreases, the approximations f^j turn coarser with values on intervals of half the size as those for V_{j+1} (see gray-colored basis functions in Fig. 4.4). Eventually, the so-called *reference space* V_0 is reached, containing step functions on unit intervals. The overall sequence of spaces $(V_j)_{j=0,\ldots,J}$ fulfills

$$V_0 \subset V_1 \subset V_2 \subset \ldots$$

where V_0 can be build by the set of integer-shifted versions of the Haar scaling function $\{\phi_{0,k} = \phi(\cdot - k), k \in \mathbb{Z}\}$. As obvious from Fig. 4.4, an orthonormal basis for any other

[3] A sequence f is square-summable if $\sum_{x\in\mathbb{Z}} f^2(x) < \infty$.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

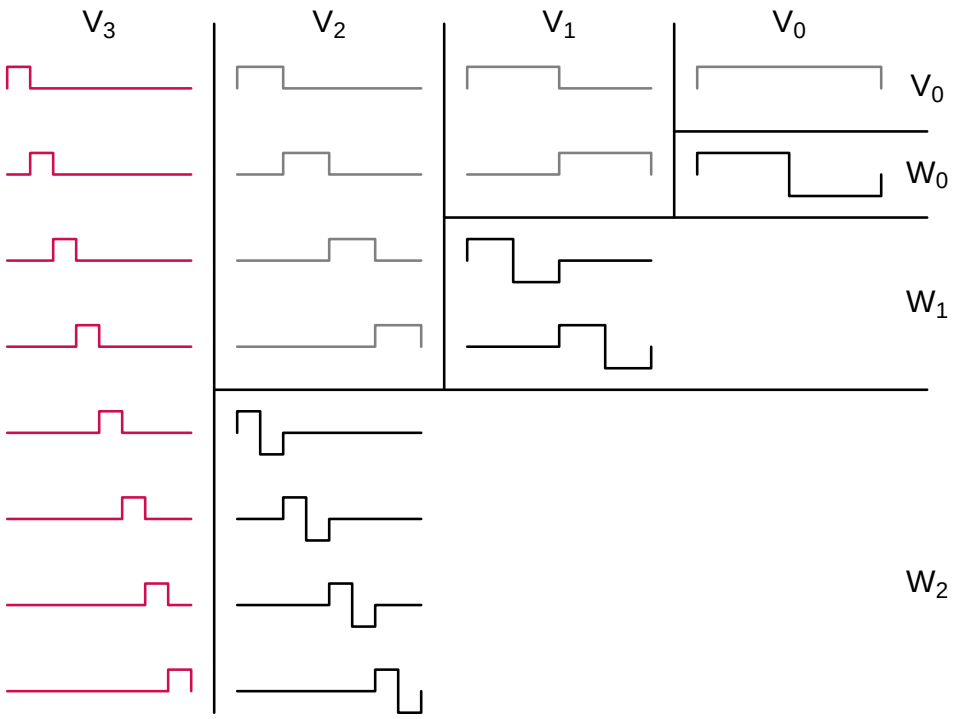


Figure 4.4: Repeated splitting of a vector space into two orthogonal subspaces build the basis for the discrete wavelet transform.

$V_j, j > 0$, is formally obtained by dilations of the father wavelet: $\{\phi_{j,k} = 2^{j/2}\phi(2^j \cdot -k), k \in \mathbb{Z}\}$. As a consequence, the central *two-scale relationship* can be established:

$$\phi_{j,k}(x) \quad = \quad \big(\phi_{j+1,2k}(x) + \phi_{j+1,2k+1}(x)\big)/\sqrt{2}.$$

This means that a function at a given level evolves by averaging at the next higher level (see also Figure 4.4).

Whenever a function f^j is expressed in terms of the next coarser approximation, the residuals represent the details of f^j that are lost due to the approximation. In this case, the details are constants of the same resolution as the function f^j itself and can thus be written as linear combination of the wavelets $\psi_{j-1,k}, k \in \mathbb{Z}$. As obvious from Figure 4.3, two wavelets with the same dilation index j but differing k cannot have overlapping support. Intra-level orthogonality is therefore guaranteed, so that the set of wavelets at a fixed level j defines a *detail space*:

$$W_j \quad = \quad \text{span}\{\psi_{j,k}, k \in \mathbb{Z}\}.$$

In the Haar example, it is easily ascertained that $\phi_{j,k}$ and $\psi_{j',k'}$ are orthogonal for $j \leq j'$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

(compare black-colored functions in Figure 4.4). This implies that W_{j-1} is the orthogonal complement of V_{j-1} in V_j:

$$V_j \;=\; V_{j-1} \oplus W_{j-1}. \tag{4.4}$$

Similar to the two-scale equation, the relationship of Haar father (Eq. 4.1) and mother wavelet (Eq. 4.2) can be generalized:

$$\psi_{j,k}(x) \;=\; \big(\phi_{j+1,2k}(x) - \phi_{j+1,2k+1}(x)\big)/\sqrt{2}.$$

Hence, the residuals at a fixed level arise from differencing a function at the next higher level.

Now, mutual orthogonality across levels holds since one of the wavelets is constant in the shared support region (compare Figure 4.3). With this inter-space orthogonality between W_j and $W_{j'}, j \neq j'$, recursive decomposition in a global approximative part and level specific details yields

$$V_j \;=\; V_{j-1} \oplus W_{j-1} = V_{j-2} \oplus W_{j-2} \oplus W_{j-1} = \ldots = V_0 \oplus \bigoplus_{l=0}^{j-1} W_l. \tag{4.5}$$

To be more general, replace V_0 with V_{j_0} in Eq. (4.5), where $j_0 < j$ denotes the primary decomposition level. It furthermore holds that $L^2(\mathbb{R}) = \overline{\bigcup_{j\in\mathbb{Z}} V_j} = \overline{V_0 \oplus \bigoplus_{j=0}^{\infty} W_j} = \overline{\bigoplus_{j\in\mathbb{Z}} W_j}$ (Härdle et al., 1998).

In conclusion, a finite and noisy signal of dyadic length, $x \in \{1, \ldots, n = 2^J\} \subset \mathbb{Z}$, has the following wavelet expansion:

$$\begin{aligned} y(x) &= f(x) + \varepsilon(x) \\ &= \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{J-1}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(x) + \varepsilon(x), \quad \varepsilon(x) \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2). \end{aligned} \tag{4.6}$$

The choice of the initial dilation level j_0 determines the quality of the approximation (Ogden, 1997), but will depend on personal experience and the given data because of missing binding guidelines. If $j_0 > 0$, one speaks of *partial* decomposition and, if $j_0 = 0$, of *complete* decomposition. In general, the number of vanishing moments of the father wavelet determines the degree to which a polynomial is still exactly reproducible in V_0. Hence, V_0 is equivalent to the null space of the smoothing spline penalty (Hastie et al.,

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2001, Chap. 5.9).

Provided the recorded time series is scaled to the unit support interval, the complete decomposition involves as many wavelet coefficients as there are observations. Namely, there is the father plus 2^j contributions at each level $j = 0, \ldots, J-1$. Summation of the geometric sequence leads to $1 + \sum_{j=0}^{J-1} 2^j = 1 + (2^J - 1) = n$. Note that the coefficient $\alpha_{0,0}$ coincides with the sample mean in the Haar case. Since an increase in level entails a bisection of the wavelet's support of 2^{-j} on simultaneous doubling of the frequency (Figure 4.3), the *resolution* enhances accordingly. Attention should be paid to the notation. Some authors revert the j-axis so that father and mother do no more lie at level $j = 0$. Level j is then rather used in the meaning of decomposition depth. The definition in Eq. 4.3 changes accordingly. In addition, the technical term *scale* sometimes replaces *level* or two to the power of *level*.

(Inverse) discrete wavelet transform

The totality of the coefficients α_{0k} and β_{jk} is called *discrete wavelet transform* (DWT). It is determined by the projection of the function f onto the subspaces V_0 and $W_j, j = 0, \ldots, J-1$, and reflects the degree of congruity of the function with the wavelets ϕ_{0k} and ψ_{jk}, respectively. While the α_{0k}'s summarize the general form of the function, the β_{jk}'s represent local details. In contrast to the Fourier transform, that operates on a single resolution level, the wavelet transform works on multiple resolution levels (Figure 4.1). The multiresolution property constitutes the ability of the wavelet transform to capture nonstationary events and to map irregulary occurring features. Hence, the recursive decomposition of an arbitrary function in $L^2(\mathbb{R})$ in its smooth and rough parts is also termed *multiscale* or *multiresolution analysis*. This becomes also immediately plausible from the example given in Fig. 4.4. A formal definition can be found in e. g. Ogden (1997). Metaphorically speaking, the visual acuity reduces continuously while stepping forward in the hierachically organized decomposition. For instance, leaving a forest can be compared to going away from fine structures of tiny leaves over twigs and branches, passing by stems and whole trees to the entire wood till mere green color remains. The peculiarity of the wavelet transform arises from the concurrence of the fine and coarse patterns.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Giving an example, Figure 4.5 depicts the decomposition of an NMR signal of length $n = 1024$ (Wavelab[4]) on the left side. The vertical bars correspond by position to the location and by length to the quantity of the level specific wavelet contributions. The largest weights emerge from slowly changing features, whereas highly resolved features do not play a big role for the representation of the NMR signal. Solely, the exclusive hallmark of the sharp spike is correctly caught by a few high frequency wavelets that have small support. This ideally demonstrates the adaptivity of wavelet analysis.

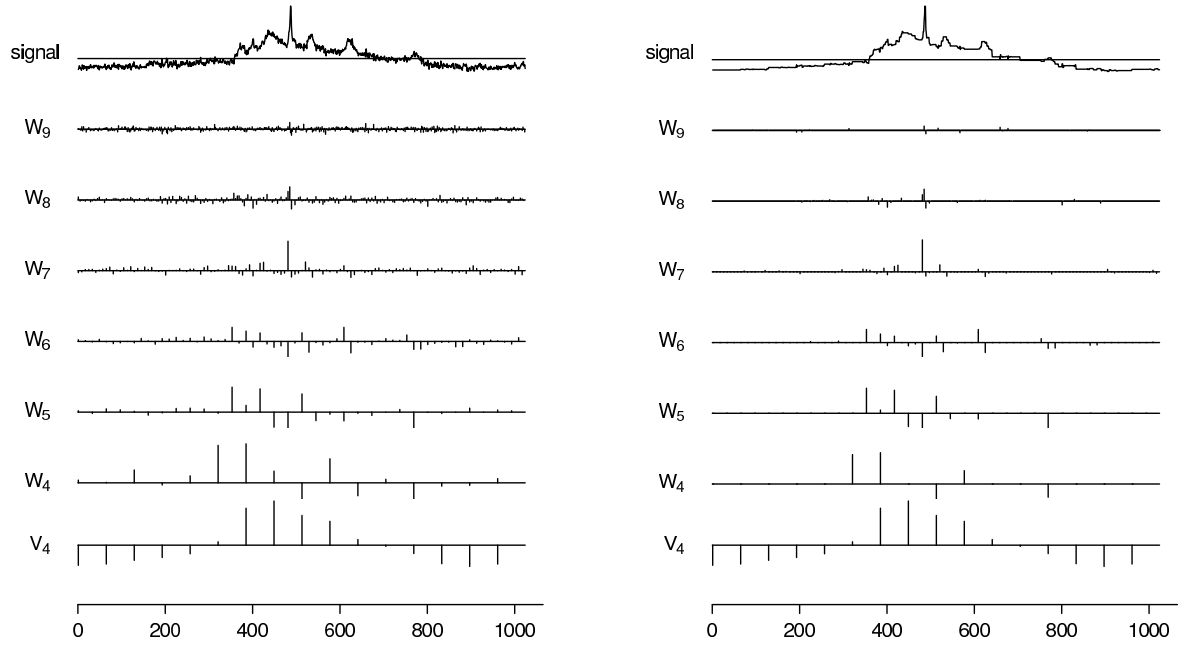


Figure 4.5: Decomposition (left) and denoised reconstruction (right) of an NMR signal using Haar wavelets.

[4]Wavelab is a free library of Matlab rountines for wavelet analysis, downloadable from http://www-stat.stanford.edu/∼wavelab

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

To gain further insight into the properties of wavelet decomposition and to introduce the inverse operation, Eq. (4.6) is reformulated in matrix notation setting $j_0 = 0$:

$$\begin{aligned}
\begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(n) \end{pmatrix} &= \begin{pmatrix} \phi_{00}(1) & \psi_{00}(1) & \psi_{10}(1) & \psi_{11}(1) & \cdots & \psi_{J-1,\frac{n}{2}-1}(1) \\ \phi_{00}(2) & \psi_{00}(2) & \psi_{10}(2) & \psi_{11}(2) & \cdots & \psi_{J-1,\frac{n}{2}-1}(2) \\ \vdots & & & & & \\ \phi_{00}(n) & \psi_{00}(n) & \psi_{10}(n) & \psi_{11}(n) & \cdots & \psi_{J-1,\frac{n}{2}-1}(n) \end{pmatrix} \begin{pmatrix} \alpha_{00} \\ \beta_{00} \\ \beta_{10} \\ \beta_{11} \\ \vdots \\ \beta_{J-1,\frac{n}{2}-1} \end{pmatrix} \\
&= \mathbf{W}\boldsymbol{\gamma}
\end{aligned}$$

Solving the corresponding system of normal equations leads to the ordinary least squares estimate of the wavelet transform:

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} &= (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} \\
&= \mathbf{W}'\mathbf{y},
\end{aligned}$$

the latter equality being due to the orthogonality of the basis functions in $\mathbf{W}$. In fact, this decomposition is loss-free, meaning that the information transported by $\mathbf{y}$ entirely passes into the wavelet domain. Although concentrated in a much smaller fraction of the components than in the raw data, the wavelet transform maintains the sample variance, i. e. is *energy (variance) preserving*:

$$\mathrm{Var}(\hat{\boldsymbol{\gamma}}) = \mathbf{W}' \,\mathrm{Var}(\mathbf{y})\mathbf{W} = \sigma^2\mathbf{I} = \mathrm{Var}(\mathbf{y}).$$

Even though, in the more general case, $\mathbf{y}$ might be correlated data with $\mathrm{Var}(\mathbf{y}) = \boldsymbol{\Sigma}$, the DWT has the ability to dissolve the correlation structure and to produce a sparse covariance matrix $\mathrm{Var}(\hat{\boldsymbol{\gamma}}) = \mathbf{W}'\boldsymbol{\Sigma}\mathbf{W}$. This means the wavelet coefficients exhibit a flat, almost white spectrum (Gençay et al., 2002, Chap. 5). Hence, the discrete wavelet transform is esteemed for its decorrelation or whitening property. The fMRI community recently detected this fundamental feature for their purpose of multiple test-correction: Significant clusters of activated voxels can be determined in the wavelet domain by the simple Bonferroni correction (Van de Ville et al., 2006).

Define the *inverse discrete wavelet transform* (IDWT) to be

$$\hat{\mathbf{y}} = \mathbf{W}\hat{\boldsymbol{\gamma}}.$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Again supposing $j_0 = 0$, even perfect reconstruction of the input signal is accomplished:

$$\hat{\mathbf{y}} \quad = \quad \mathbf{W}\hat{\gamma} = \mathbf{W}\mathbf{W}'\mathbf{y} = \mathbf{W}'\mathbf{W}\mathbf{y} = \mathbf{y}.$$

Here, 'perfect' is used in the sense of completely recoverable as no information is lost.

Preview on denoised reconstruction

Yet, the analyst's interest does usually not focus on the recorded measurements but on strategies to reveal the underlying function. As anticipated previously, the wavelet decomposition allows to determine basis functions of marginal meaning by their absolute weight. The left plot of Figure 4.5 suggests to discard, above all, coefficients at large levels where extremely localized features of high frequency are mapped. Similar to analysis of variance, coefficients are considered to be noise or, on the other hand, significantly important according to their explanatory strength. Applying an appropriate cut-off value and performing the inverse wavelet transform yields a regularized version of the initial NMR signal. The result for the Haar basis is depicted on the right side of Figure 4.5. Particularly coefficients at large levels vanish, thereby suppressing rapid changes of the input signal. Though the essential form is clearly worked out, smoother wavelets are obviously required to meet the nature of the true NMR signal. On the other hand, the characteristic sharp peak is appropriately reproduced since wavelets concentrate the mass of oscillations on a small interval. Note the single distinct deflexion in W_7 that indicates the singularity of the original series. In total, a sparse and compressed wavelet representation of the input data is achieved. This mechanism also underlies the standardized image format JPEG2000[5] which allows the user to specify a compression rate. If no compression is desired, the loss-free reconstruction case is given with the same memory amount as required by the object itself. In the compression case, the use of special decoders leads to considerable memory reduction storing non-zero coefficients only. Multidimensional wavelets are presented in Section 4.1.3, while Section 4.1.4 deals with different *thresholding* and *shrinkage* approaches.

[5]JPEG is based on cosine functions; JPEG2000 works with wavelets.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Practical matters

In view of the above, two more practical issues come to mind. First, the Haar basis clearly adapts to the unit interval, any finite sequence can be scaled to. But wavelets of other families cannot straightforwardly be forced to have support $\subseteq [0,1)$. As a consequence, discontinuities at the interval boundaries can occur when a multiresolution analysis of the sequence is performed. These might be mitigated by symmetric or periodic expansion of the signal. Second, if a non-dyadic sampling rate is encountered, zero-padding or interpolating of the signal up to dyadic length has to be considered beside symmetric and periodic boundary handling (Ogden, 1997). Some approaches will introduce correlations among the coefficients. Rather than manipulating the data, other transforms for signals of unconstrained length seem advisable such as the maximum-overlap discrete wavelet transform (MODWT) (Gençay et al., 2002), biorthogonal wavelet transforms (Ogden, 1997), or lifting schemes (Jansen, 2001). The MODWT is also known as non-decimated, redundant, translation-invariant, or stationary wavelet transform.

The fast wavelet transform algorithm

When it comes to implementation of the (inverse) discrete wavelet transform, Mallat's pyramidal algorithm is usually employed for its computational efficiency. The core aspect here is that the decomposition in smooth and rough parts can be interpreted as application of high- and low-pass filters. Indeed, the low-pass filter coefficients conform with the two-scale relationship of the scaling function:

$$l_k \;=\; <\phi, \phi_{1,k}> \overset{\text{Haar}}{=} \begin{cases} \frac{1}{\sqrt{2}}, & k = 0,1 \\ 0, & \text{otherwise.} \end{cases}$$

By m-times application of filter operator L, represented by the sequence $\{l_k\}_{k\in\mathbb{Z}}$, to the scaling coefficients at a particular level, represented by $\alpha_{j,\cdot} = \{\alpha_{j,k}\}_{k\in\mathbb{Z}}$, scaling function coefficients at the m-times lower level are obtained:

$$\alpha_{j-m,\cdot} \;=\; L^m \alpha_{j,\cdot} = L^{m-1} \left\{ \sum_{i\in\mathbb{Z}} l_{i-2k} \alpha_{j,i} \right\}_{k\in\mathbb{Z}} .$$

The second equality sheds some light on the downsampling by two, inherent to each filtering step (compare Figure 4.6). This is hidden in the indices of $\{l_i\}$ that are all offset

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

by two.

The high-pass filter operator H can be defined by the quadratur mirror relationship (see e. g. Gençay et al., 2002):

$$h_k = (-1)^k l_{1-k}, \quad k \in \mathbb{Z}.$$

Similar to the relation of father and mother wavelet, unique application of H to the scaling coefficients at level j leads to the wavelets coefficients at level $j-1$:

$$\beta_{j-1,\cdot} = H\alpha_{j,\cdot}.$$

Combining L and H yields

$$\beta_{j-m,\cdot} = HL^{m-1}\alpha_{j,\cdot}.$$

Iteratively subsampling every other coefficient in either filter steps gives the pyramidal or cascade scheme of the decomposition illustrated in Figure 4.6. The reconstruction follows

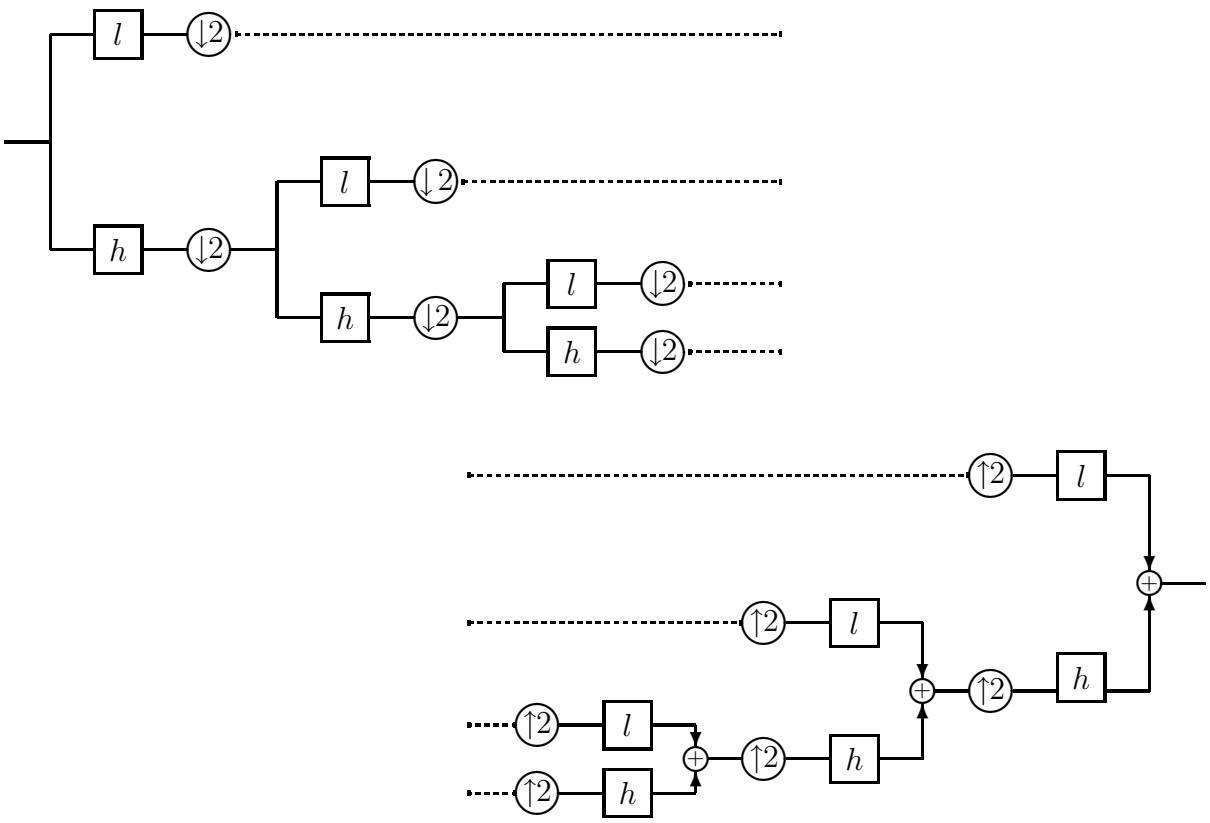


Figure 4.6: Decomposition (top) and synthesis (bottom) of a signal using the pyramidal algorithm. The branches refer to convolution with either low-pass ($\boxed{l}$) or high-pass filter coefficients ($\boxed{h}$). The icons $\downarrow 2$ and $\uparrow 2$ denote down- and upsampling by two.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

the reverse principle. At a particular level, the upsampling step performs zero-padding and is necessary for the convolution with the filter sequence of double length.

The combination of low- and high-pass filters results in band-pass filters. Specifically, a band-pass filter over the frequency interval $[1/4, 1/2]$ results from serial application of an ideal filter pair L and H. For this purpose let L pass the frequencies lower than $1/2$ and H capture frequency dynamics greater than $1/4$. Therefore, one says that the pryamidal algorithm decomposes an input signal into sub-bands. In addtion, this viewpoint allows the following conclusion: The closer a wavelet filter approaches an ideal filter the less leakage of frequencies occurs, thus, the better the edge preservation property becomes (Gençay et al., 2002). Beyond this, the primary decomposition level is often chosen on a desired low-pass cut-off frequency.

4.1.3 Multidimensional wavelets decomposition - a 2d example

One-dimensional wavelets can easily be extended to the higher dimensional case via tensor products (see e. g. Ogden, 1997). Note that consecutive operations on columns, rows, layers etc of n-dimensional observational arrays generate *separable* basis functions. It is well known that these functions are likely to lack isotropy. In contrast, *non-separable* wavelets offer more freedom but at the expense of computational cost. Therefore, the discussion limits itself to separable wavelet functions. The decomposition idea is exemplified by means of 2d wavelets intended to analyze square-integrable 2d functions $f(x, y) \in L^2(\mathbb{R}^2)$.

Wavelet basis functions in $\mathbb{R}^2$

To begin with, define the 2d father: $\phi(x, y) = \phi(x)\phi(y)$. As before, translations and dilations of the father,

$$\{\phi_{j,k_1,k_2}(x, y) = 2^j\phi(2^jx - k_1, 2^jy - k_2), j, k_1, k_2 \in \mathbb{Z}\}$$

constitute a basis for $V_j^{x,y} = V_j^x \otimes V_j^y$. Deploying the multiresolution property and Eq. (4.4), the space $V_{j+1}^{x,y}$ of 2d block functions with resolution $2^{j+1} \times 2^{j+1}$ can be approximated by the successive space $V_j^{x,y}$ of twice as coarse squares and the complimentary

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

residual space:

$$\begin{aligned} V_{j+1}^{x,y} &= V_{j+1}^{x} \otimes V_{j+1}^{y} \\ &= (V_j^x \oplus W_j^x) \otimes (V_j^y \oplus W_j^y) \\ &= (V_j^x \otimes V_j^y) \oplus \big((V_j^x \otimes W_j^y) \oplus (W_j^x \otimes V_j^y) \oplus (W_j^x \otimes W_j^y)\big) \\ &= V_j^{x,y} \oplus W_j^{x,y}. \end{aligned}$$

The rewriting reveals that the 2d space $W_j^{x,y}$ is made up of three orthogonal subspaces, each of which endowed with a tensor product basis of its components. So define the corresponding 2d mother wavelets

$$\begin{aligned} \psi^1(x,y) &= \phi(x)\psi(y) \\ \psi^2(x,y) &= \psi(x)\phi(y) \\ \psi^3(x,y) &= \psi(x)\psi(y) \end{aligned}$$

and build the wavelet basis for $W_j^{x,y}$ by means of translation and dilation:

$$\{\psi_{j,k_1,k_2}^d(x,y) = 2^j\psi^d(2^jx - k_1, 2^jy - k_2), d = 1,2,3, j, k_1, k_2 \in \mathbb{Z}\}.$$

Note that the basis functions are tensor products within the same scale. This distinguishes the *square* wavelet transform from the rectangular wavelet transform that involves tensor product across scales (Jansen, 2001). Figure 4.7 shows the four functions $\phi(x,y), \psi^1(x,y), \psi^2(x,y)$ and $\psi^3(x,y)$ associated with the 1d symlet with two vanishing moments. The clearly recognizable orientation in space tells us that ψ^1 emphazises edges in the horizontal direction, ψ^2 in the vertical and ψ^3 in the diagonal direction. These properties can be understood in view of the low- and high-pass filter analogons and are most intuitively accessed through the shapes of 2d Haar wavelets (Figure 4.8). Note also that the effective support of 2d tensor product wavelets shrinks by 2×2 when increasing the dilation parameter by one.

A finite and noisy 2d signal, e. g. an $(n_1 \times n_2)$-image Y, thus has the following wavelet representation:

$$\begin{aligned} Y(x,y) &= f(x,y) + \varepsilon(x,y) \\ &= \sum_{k_1=0}^{2^{j_0}-1}\sum_{k_2=0}^{2^{j_0}-1} \alpha_{j_0,k_1,k_2}\phi_{j_0,k_1,k_2}(x,y) + \sum_{d=1}^{3}\sum_{j=j_0}^{J-1}\sum_{k_1=0}^{2^j-1}\sum_{k_2=0}^{2^j-1} \beta_{j,k_1,k_2}^d \psi_{j,k_1,k_2}^d(x,y) + \varepsilon(x,y), \end{aligned}$$

with $\varepsilon(x,y) \sim N(0,\sigma^2)$, $J = \log_2(\min(n_1,n_2))$.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

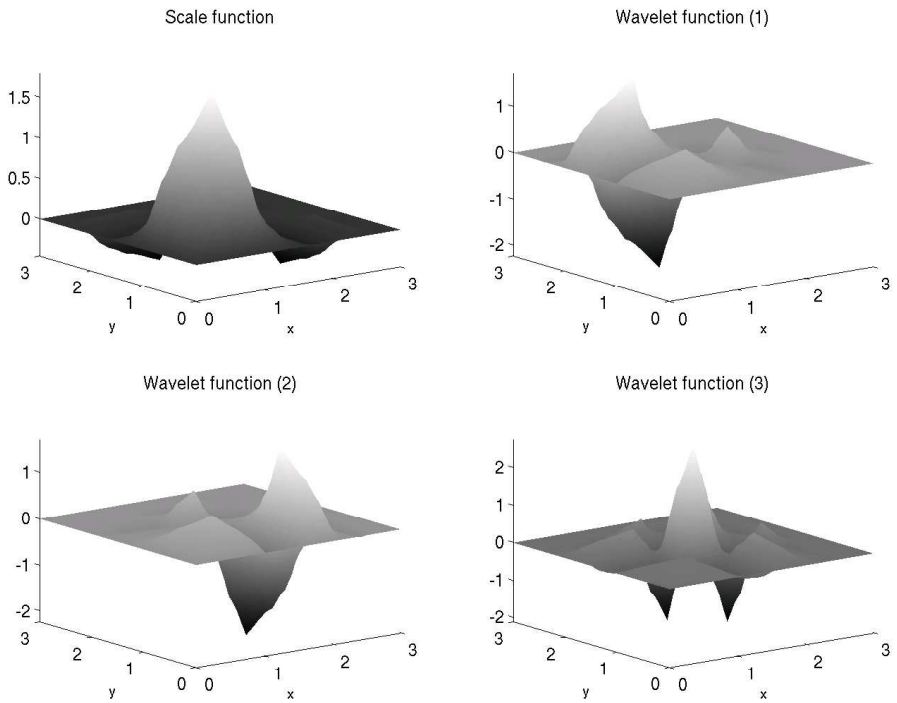


Figure 4.7: Scaling function and mother wavelet functions of 2d symlet with two vanishing moments.

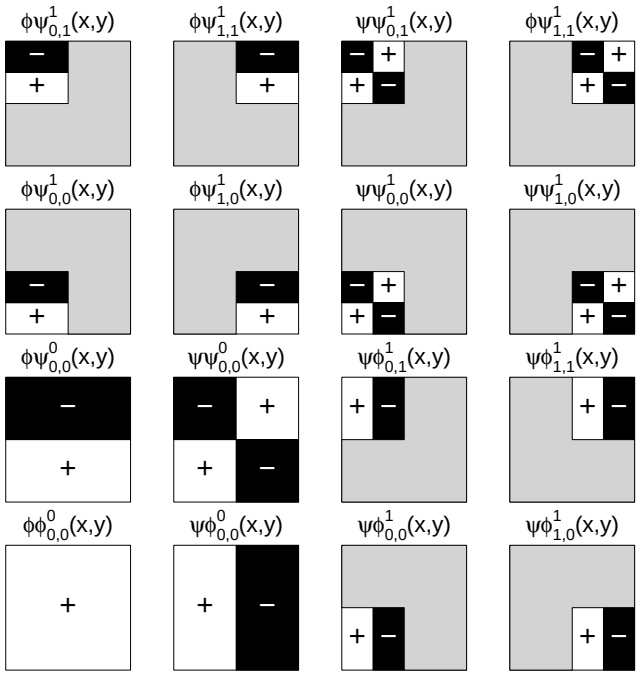


Figure 4.8: Exemplary basis functions of the 2d Haar family, projected onto the xy-plane. The exponent denotes the level j, while the subscripts correspond to shift parameters k_1 and k_2.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2d decomposition scheme

In analogy to the 1d case, the pyramidal algorithm also adapts to the 2d wavelet decomposition. In a first step, a pair of low- and high-pass filters runs over the rows, and afterwards along the columns of the given image. This procedure is repeated on each purely low-pass filtered output, i. e. the coefficient matrix $(\alpha_{j,k_1,k_2})_{\substack{k_1=0,\dots 2^j-1 \\ k_2=0,\dots 2^j-1}}$. It is denoted by LL_j in Figure 4.9 for pointing out the filter link and filter order. Note that LL_J encodes for the input image itself. Convolution of the detail wavelets $\psi^1(x,y), \psi^2(x,y)$ and $\psi^3(x,y)$ with the approximating part LL_{j+1} results in horizontal, vertical and diagonal detail components LH_j,HL_j, and HH_j, respectively. Each iteration causes the filtered output to be downsampled by 2 per direction which establishes the pyramidal structure and allows for the nested visualization. It also gives the algorithm its speed since in each step only 1/4 of the number of coefficients needs to be decomposed. The inverse process of reconstruction is governed by dyadic upsampling.

A black-and-white picture of the Siegestor in Munich serves as subject in Figure 4.10. Juxtaposed to the (256×256)-image matrix are two nested steps of the 2d DWT. The

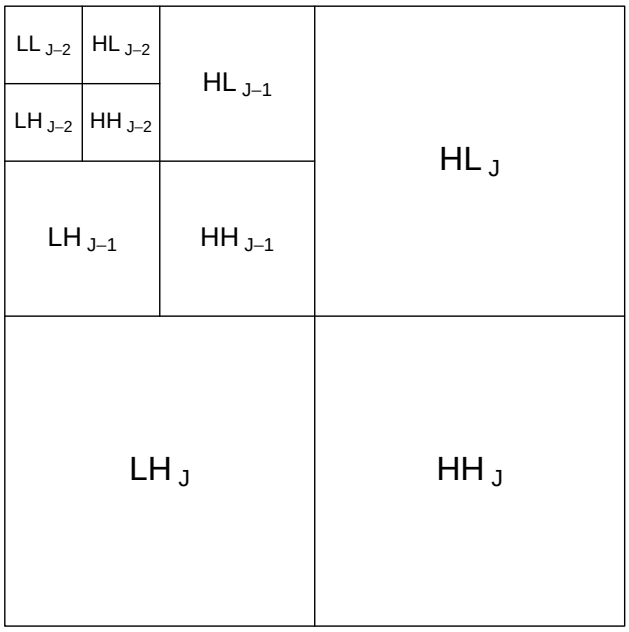


Figure 4.9: Nested two-dimensional signal decomposition on three scales.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Figure 4.10: Original greyscaled image and 2d DWT based on Haar wavelets.

decomposition units exhibit their directional preferences in plain manner: brightness contrasts on columns (vertical features) are extracted separately from beams (horizontal features) and perspective alignments (diagonal features). It is worth emphazising that components including a high-pass filtering step have been adjusted for comparable visualization with the residual image. Background sprinkles are present, though hardly visible to the human eye due to poor reproduction quality. They can however be removed in an intermediate denoising, i. e. thresholding, step before performing the 2d IDWT.

By means of the Siegestor photography and the 2d Haar basis, I would like to elucidate the successive approximations associated with spaces $V_J \supset \ldots \supset V_0$. Figure 4.11 demonstrates how with decreasing level the grey colores contained in a scaling image concentrate on a "grand mean" value. At the same time, the wavelet's blocky support reaches the size of the input image. Subsequent to the third decomposition level, the coefficients do no more reflect any visual features of the original image. Therefore, an appropriate DWT would stop at this specific level, just allowing to sort out features of reasonable frequency. In particular, variance reduces rapidly during the last four decomposition levels. Hence, a 'kink' in the curve of the squared sum of scaling coefficients could indicate an adequate primary decomposition level j_0.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

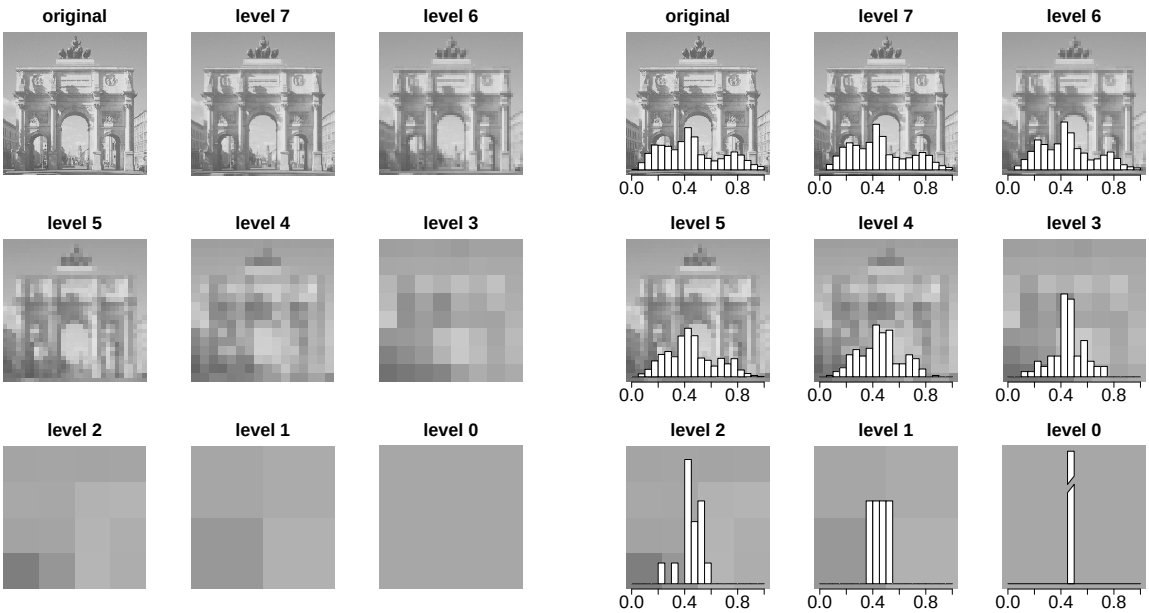


Figure 4.11: Purely low-pass filtered coefficient matrices of different levels (left) with overlaid normalized histograms (right).

Generalization to 3d

With the thorough introduction and examination of 1d and 2d wavelets as well as of the corresponding decomposition schemes, the extension to the 3d wavelet analysis is straightforward. In brief, the 3d scaling function $\phi(x, y, z)$ and seven wavelets in all cross-spatial horizontal, vertical and diagonal directions are designed as tensor product of 1d wavelets:

$$
\begin{aligned}
\phi(x,y,z) &= \phi(x)\phi(y)\phi(z) \\
\psi^1(x,y,z) &= \phi(x)\phi(y)\psi(z) \\
\psi^2(x,y,z) &= \phi(x)\psi(y)\phi(z) \\
\psi^3(x,y,z) &= \psi(x)\phi(y)\phi(z) \\
\psi^4(x,y,z) &= \psi(x)\psi(y)\phi(z) \\
\psi^5(x,y,z) &= \psi(x)\phi(y)\psi(z) \\
\psi^6(x,y,z) &= \phi(x)\psi(y)\psi(z) \\
\psi^7(x,y,z) &= \psi(x)\psi(y)\psi(z).
\end{aligned}
$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

All operations, such as dilation and translation, orthogonal subspace representation and the pyramidal filterbank algorithm, can be adapted to the above 3d initial wavelet functions. Finally, the wavelet analysis of square-summable function $f(x, y, z) \in l^2(\mathbb{R}^3)$ can be conducted.

4.1.4 Thresholding and shrinkage

Given the motivating preview on denoising (p. 97), this section serves to state more precisely what is meant by removing "small" coefficients. Let me derive noise reduction procedures from the commonly used additive noise model. To ease readability, the location parameter is dropped for the moment. Suppose the observations $\mathbf{y}$, $\mathbf{y} \in \mathbb{R}^n$, relate to the true function f but are corrupted by Gaussian white noise, i. e. the observations are realizations of independent and identically distributed Gaussian random variables:

$$\mathbf{y} \;=\; f + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}).$$

Due to the linearity of the wavelet transform, the additivity remains unchanged:

$$\begin{array}{ccccc} \underbrace{\mathbf{W}'\mathbf{y}} & = & \underbrace{\mathbf{W}'f} & + & \underbrace{\mathbf{W}'\boldsymbol{\varepsilon}} \\ \hat{\boldsymbol{\gamma}} & = & \boldsymbol{\gamma} & + & \boldsymbol{\eta}\,. \end{array}$$

Obviously, the wavelet coefficients of a noisy sample can be regarded as noise contaminated observations of hidden wavelet coefficients. Moreover, if $\mathbf{W}$ is orthogonal, the wavelet transform of the stationary and white noise will also be stationary and white:

$$\mathrm{Var}(\boldsymbol{\eta}) = \mathrm{Var}(\mathbf{W}'\boldsymbol{\varepsilon}) = \mathbf{W}' \,\mathrm{Var}(\boldsymbol{\varepsilon})\mathbf{W} = \sigma^2 \mathbf{I} = \mathrm{Var}(\boldsymbol{\varepsilon}). \tag{4.7}$$

The wavelet transform creates a sparse signal representation in the sense that most empirical coefficients are zero or close to zero (compare Fig. 4.5). Only a few distinct coefficients reflect important signal singularities beyond the overall shape. In other words, the transform preserves the sum of squares (isometry of risks), but concentrated in a much smaller fraction of the components than in the raw data.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Thresholding rules

The most straightforward method for noise reduction suggests to let the informative coefficients survive while rigorously zeroing every "small" coefficient. This "keep-or-kill" strategy involves a well-chosen threshold λ that demarcates the cutoff between essential variables (or basis functions) and noise. The method is therefore also called *hard thresholding* and can formally be expressed as a function δ:

$$\hat{\gamma}_{\lambda,i} \quad = \quad \delta^H(\hat{\gamma}_i, \lambda) = \begin{cases} \hat{\gamma}_i, & |\hat{\gamma}_i| \geq \lambda \\ 0, & \text{otherwise.} \end{cases}$$

Figure 4.12 displays the corresponding function on the left hand side with $\lambda = 1$.

The second well-known denoising approach is *soft thresholding*:

$$\hat{\gamma}_{\lambda,i} \quad = \quad \delta^S(\hat{\gamma}_i, \lambda) = \text{sign}(\hat{\gamma}_i)\,(|\hat{\gamma}_i| - \lambda)_+ = \begin{cases} \hat{\gamma}_i + \lambda, & \hat{\gamma}_i < -\lambda \\ 0, & -\lambda \leq \hat{\gamma}_i < \lambda \\ \hat{\gamma}_i - \lambda, & \hat{\gamma}_i \geq \lambda\,. \end{cases}$$

The difference to hard thresholding is that, in addition to the truncation at λ, all other wavelet weights are shrunken by λ. This way soft thresholding becomes a continuous operation (right side of Fig. 4.12), offering considerable advantages. For instance, noise coefficients that have spuriously passed the hard threshold may appear as annoying 'blips'.

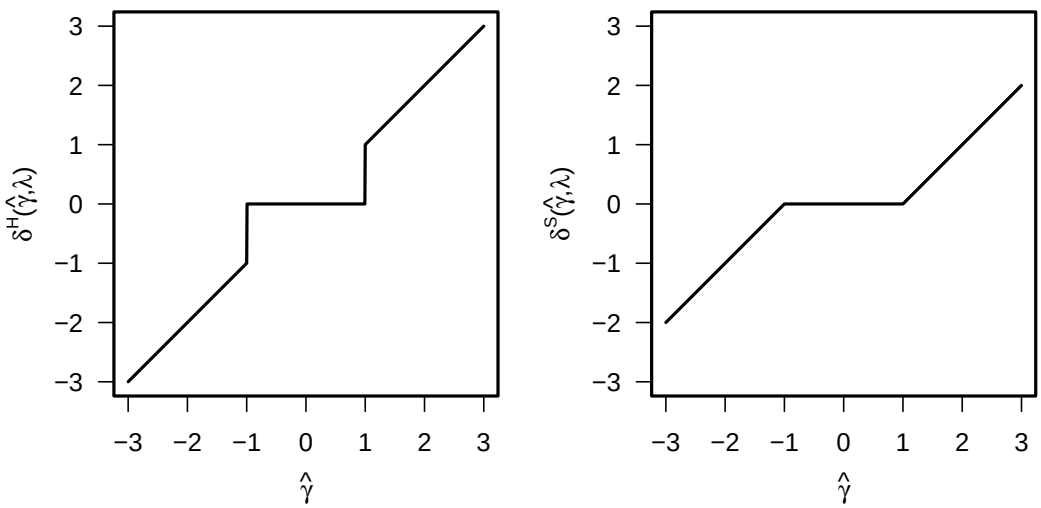


Figure 4.12: Classical hard (left) and soft (right) thresholding rules with $\lambda = 1$.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Yet, soft thresholding shrinks these false structures and does therefore not suffer as much from high variability (Hastie et al., 2001). Soft thresholding can be termed a "high bias – low variance" solution to the shrinkage problem, whereas hard thresholding is a "low bias – high variance" solution.

The denoising mechanisms can be embedded into the general class of non-parametric regression models. Aiming at regularization, the objective to be minimized is formulated as trade-off between closeness to the data and sparsity/smoothness, governed by a tuning parameter:

$$\|\mathbf{y} - \mathbf{W}\boldsymbol{\gamma}_\lambda\|_2^2 + \lambda\,\mathrm{Pen}(\boldsymbol{\gamma}_\lambda) \quad \longrightarrow \quad \min_{\boldsymbol{\gamma}_\lambda} . \tag{4.8}$$

In the signal processing language, this balance between fidelity to the data and smoothness penalty is described as compromise between 'energy' and 'entropy'. Energy is an expression of signal size (compare p. 89) and related to variance. The entropy is a measure of the disorder of a system and is linked to the information content of that system. On the one hand, one could say that a high entropy state is information poor because there is so much disorder, and the disorder is essentially random. On the other hand, one could argue that a high entropy state is information rich because a truthful description of the exact state of randomness in all its detail would require lots of information. It is this latter sense which underlies Shannon's definition for discrete data: The information entropy is the log of the number of accessible states, and is dimensionless.Attention has to be paid to avoid possible confusion because this definition is a measure of the *lack* of order in a system. In Fig. 4.11, the entropy of the scaling image decreases with level until one single gray value remains. This monochromic state contains the least information according to Shannon's entropy. In contrast, a uniformly distributed histogram would have maximum entropy and maximum information.

To proceed with the topic of smoothing approaches, recapitulate that $\mathbf{y} = \mathbf{W}\hat{\boldsymbol{\gamma}}$ by perfect reconstruction. Then the orthogonality of the wavelet transform allows to replace the data term by the squared differences between untouched coefficients and thresholded counterparts:

$$\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_\lambda\|_2^2 + \lambda\,\mathrm{Pen}(\boldsymbol{\gamma}_\lambda) \quad \longrightarrow \quad \min_{\boldsymbol{\gamma}_\lambda} .$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

The penalty term can take on different forms. For example, using the L_2 norm

$$\text{Pen}(\boldsymbol{\gamma}_\lambda) \quad = \quad \|\boldsymbol{\gamma}_\lambda\|_2^2 = \sum_i \gamma_{\lambda,i}^2,$$

yields a scaled version of the ordinary least squares coefficients:

$$\hat{\boldsymbol{\gamma}}_\lambda \quad = \quad \frac{1}{1+\lambda}\hat{\boldsymbol{\gamma}}.$$

Unless the intercept, i. e. the scaling wavelet coefficients α_{0k}, is left out from penalization, this corresponds to *linear shrinkage* of the observed data:

$$\hat{\mathbf{y}} \quad = \quad \frac{1}{1+\lambda}\mathbf{y}.$$

The procedure is also known as ridge regression (Hastie et al., 2001, Chap. 3). However, the L_2 norm measures energy rather than sparsity or smoothness: selection does not take place. Hence consider the use of the L_1 norm

$$\text{Pen}(\boldsymbol{\gamma}_\lambda) \quad = \quad \|\boldsymbol{\gamma}_\lambda\|_1^2 = \sum_i |\gamma_{\lambda,i}|,$$

which leads to the same framework as in Eq. (4.8), also familiar under the name of lasso regression (Hastie et al., 2001, Chap. 3). Because $\mathbf{W}$ is orthogonal, the lasso criterion results in the simple solution of soft thresholding. The method is highly esteemed for its adaptivity founded on least absolute shrinkage *and* selection from the parameter space. A wavelet fit constrained to the L_1 norm represents a *non-linear* smoothing procedure. Likewise, hard thresholding turns out to be a non-linear operation. In this case, the objective consists of the usual residual sum of squares and a penalty term that amounts to the number of non-zero coefficients:

$$\text{Pen}(\boldsymbol{\gamma}_\lambda) \quad = \quad \sum_i 1_{[\gamma_{\lambda,i} \neq 0]},$$

where $1_{[\text{condition}]}$ is an indicator function.

Hard thresholding is said to perform better in spike detection, whereas soft thresholding is preferred if smoothness is paramount. For a discussion on risk, bias and variance of these two approaches refer to Jansen (2001, Chap. 5.5). More sophisticated rules are available such as n-degree garrote or Bayesian thresholding rules (Gençay et al., 2002; Jansen, 2001; Johnstone and Silverman, 2005). These schemes model the transition between rejected

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

and shrunken coefficients in a smoother fashion but are computationally more intensive. Note that also ridge and lasso regression can be viewed as Bayesian estimates with different priors, either Gaussian or Laplace (e. g. Hastie et al., 2001). In both models, posterior modes are derived as Bayesian estimates instead of the more commonly used posterior means. The equality of posterior mode and mean only holds in case of ridge regression. The lasso regularization constraint was promoted by Donoho and Johnstone (1995) as *SURE shrinkage* (Stein's Unbiased Risk Estimation). Donoho and Johnstone investigated optimality properties and contributed seminal work on the choice of the threshold value.

Thresholding values

Regardless of the selection of an available thresholding rule, the cut-off value λ needs to be chosen with care. If the noise proportion was known, a complete noise removal could be accomplished by hard thresholding with $\lambda = \sigma$ due to Eq. (4.7). A threshold less than the true error variance would cause undesired noise relicts while a too loose threshold leads to information loss. Yet, σ is unknown. A bunch of suitable procedures to determine λ is published, two of which will be presented here.

On the background of the additive Gaussian white noise model, a straightforward rationale is to remove all wavelet coefficients that are smaller than the expected maximum of an assumed i. i. d. normal noise sequence of given size. This approach yields the *universal threshold*:

$$\lambda^{UNIV} = \sigma\sqrt{2\log(n)}. \tag{4.9}$$

Its name reflects the global validity for all sufficiently smooth signals of length n (Jansen, 2001). It can even be stated that the universal threshold is adaptive to unknown smoothness. The interested reader may consult e. g. Härdle et al. (1998) for more detailed information on smoothness within the framework of Besov spaces.

The universal threshold operates as an oracle that reveals which coefficients are above σ but not their exact values. Donoho and Johnstone (1994) showed that, within a logarithmic factor, universal soft-thresholding performs as well as optimal wavelet coefficient selection which, in turn, is essentially not less powerful than any piecewise polynomial and spline method. Moreover, λ^{UNIV} possesses minimax property, i. e. the minimax thresh-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

old asymptotically coincides with the universal threshold. However, instead of balancing energy and entropy, smoothness comes prior to goodness of fit. This renders the results visually appealing, leading to the procedure's synonym *VisuShrink*.
The noise level in definition (4.9) has to be estimated from the data. Donoho and Johnstone (1994) use the median absolute deviation of the wavelet coefficients at the finest level $J = \log_2(n)$, divided by 0.6745. As motivated on p. 97, this level mirrors high frequency features and thus essentially noise. The use of a robust variance measure is intended to correct the bias due to some isolated signal occuring at that level. The various denoising techniques apply also to images and image volumes. Yet in the higher dimensional case, each level has multiple decomposition components, also called orientations (2d: three, 3d: seven). The question arises whether noise estimation should be based on purely high-pass filtered coefficients or, whether high- and low-pass filtered output at level J should be considered equally. Furthermore, n has to be adjusted accordingly.

If minimum risk optimality is paramount, soft-thresholding with the cut-off chosen according to Stein's Unbiased Risk Estimation (SURE) seems more suitable. Donoho and Johnstone (1995) call this procedure *SureShrink*. In contrast to the level invariant universal threshold, the SURE thresholds are independently derived from each resolution level $j = j_0, \ldots, J = \log_2(n)$ by minimizing the level specific risk $\mathrm{E}\,\|\hat{\boldsymbol{\gamma}}_{\lambda,j} - \boldsymbol{\gamma}_j\|^2$. Depending on j, the subvector $\boldsymbol{\gamma}_j = (\gamma_{j0}, \ldots \gamma_{jk}, \ldots, \gamma_{j(2^j-1)})'$ of $\boldsymbol{\gamma}$ comprises either the coefficients of the scaling functions or the ones of levelwise wavelets, here uniformly denoted by double-indexed γ_{jk}. In general, most of the wavelet literature formulates SURE shrinkage as a result in multivariate normal decision theory. Yet, I prefer to stay in the wavelet context though running the risk of complex notation.
After variance standardization in Eq. (4.7), the empirical OLS wavelet coefficients are multivariate normally distributed: $\boldsymbol{\gamma}_j^{OLS} \sim N(\boldsymbol{\gamma}_j, \mathbf{I})$. Stein proved that any (almost arbitrary) estimator for $\boldsymbol{\gamma}_j$ of the form $\hat{\boldsymbol{\gamma}}_{\lambda,j}(\boldsymbol{\gamma}_j^{OLS}) = \boldsymbol{\gamma}_j^{OLS} + g(\boldsymbol{\gamma}_j^{OLS})$ with $g : \mathbb{R}^{2^j} \longrightarrow \mathbb{R}^{2^j}$ weakly differentiable[6] leads to an unbiased estimate of the L_2 loss:

[6] Let g and h be functions defined on the real-line and integrable on every bounded interval. Then g is weakly differentiable with weak derivative h if $\int g(t)u'(t)dt = -\int h(t)u(t)dt$ is satisfied for all infinitely many times differentiable test functions u with compact support (Härdle et al., 1998). Generalization to the higher-dimensional case is straightforward.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

$$\mathrm{E}_{\boldsymbol{\gamma}_j}\|\hat{\boldsymbol{\gamma}}_{\lambda,j}\left(\boldsymbol{\gamma}_j^{OLS}\right)-\boldsymbol{\gamma}_j\|^2 \;=\; 2^j+\mathrm{E}_{\boldsymbol{\gamma}_j}\left(\|g(\boldsymbol{\gamma}_j^{OLS})\|^2+2\bigtriangledown g(\boldsymbol{\gamma}_j^{OLS})\right),$$

where $\bigtriangledown g(\boldsymbol{\gamma}_j^{OLS})=\sum_{k=0}^{2^j-1}\frac{\partial}{\partial\gamma_{jk}^{OLS}}g_k(\boldsymbol{\gamma}_j^{OLS})$. Employing soft thresholding allows the following specification:

$$g_k(\boldsymbol{\gamma}_j^{OLS})=\delta_\lambda^S(\gamma_{jk}^{OLS})-\gamma_{jk}^{OLS}=\begin{cases}\lambda, & \gamma_{jk}^{OLS}<-\lambda\\ -\gamma_{jk}^{OLS}, & -\lambda\le\gamma_{jk}^{OLS}<\lambda\\ -\lambda, & \gamma_{jk}^{OLS}\ge\lambda\,,\end{cases}$$

so that $\|g(\boldsymbol{\gamma}_j^{OLS})\|^2=\|\min(|\boldsymbol{\gamma}_j^{OLS}|,\lambda)\|^2$. Similarly, $\bigtriangledown g(\boldsymbol{\gamma}_j^{OLS})=\sum\limits_{k=0}^{2^j-1}(-1)1_{[-\lambda\le\gamma_{jk}^{OLS}\le\lambda]}=-\#\{k:|\gamma_{jk}^{OLS}|\le\lambda\}$. To sum up, Stein's estimate of risk can be rewritten as

$$SURE(\lambda,\boldsymbol{\gamma}_j^{OLS}) \;=\; 2^j-2\#\{k:|\gamma_{jk}^{OLS}|\le\lambda\}+\|\min(|\boldsymbol{\gamma}_j^{OLS}|,\lambda)\|^2.$$

Recall that the expectation $\mathrm{E}_{\boldsymbol{\gamma}_j}\,SURE(\lambda,\boldsymbol{\gamma}_j^{OLS})$ is unbiased. The all-dominant and crucial threshold is given by

$$\lambda_j^{SURE} \;=\; \arg\min_{\lambda\ge0}SURE(\lambda,\boldsymbol{\gamma}_j^{OLS}).$$

In fact, the SURE threshold takes on the value of one of the coefficients $|\gamma_{jk}^{OLS}|, k=0,\ldots,2^j-1$. Moreover, one can show that SURE shrinkage is the same as lasso regression (Hastie et al., 2001).

Risk minimization in the sense of the L_2 norm can be regarded as seeking after the equilibrium between variance and (squared) bias. Without relying on a priori knowledge about the underlying smoothness, some noise artifacts will eventually sneak in at every scale. This, however, makes the SURE shrinkage automatically smoothness-adaptive (for theoretical results see Donoho and Johnstone, 1995). One could say, that the outcome of universal thresholding is visually attractive while results from SURE shrinkage are qualitatively convincing: The reconstructions jump where the true object jumps; at the same time, the fits are smooth where the true object is smooth.

In this category of minimum risk thresholds falls also the minimax estimator itself (named *RiskShrink* by Donoho and Johnstone (1994)) and the GCV strategy (Jansen, 2001). In general, SURE and related procedures are more conservative and are more convenient when small details of f lie near the noise range. A hybrid approach was conceived in

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

order to switch from SURE to universal thresholding whenever the level specific signal-to-noise ratio is very small and the SURE result would be too noisy. Multiple alternative threshold criteria are available such as the false discovery rate (FDR) for the multiple testing problem whether the n wavelet coefficients are significantly zero (Ogden, 1997, Chap. 8.2). The FDR based threshold yields similarly smooth results as the universal threshold (Jansen, 2001; Gençay et al., 2002).

4.1.5 Resolution increment

If wavelets act as smoothing tool, interpolation does not enter the equation. If however wavelets are to be supposed to act as proper basis functions within the SVCM framework, interpolation matters. Therefore, let me briefly comment on this issue. Since, in practice, wavelet analysis uses the repeated application of filter sequences (see Section 4.1.2, pp. 98), there exists no such possibility to enhance resolution as for B-splines. This means, we cannot simply evaluate the basis functions at the desired points and sum over their estimated weight. Yet, different alternatives are conceivable.

The naive approach would be to paste multiple zeros into the level's j_0 filtering output before reconstruction. This implies a modification of the upsampling step. A more sophisticated procedure involves lifting schemes that gradually increase the complexity of the overall transform (Jansen, 2001). This is achieved by repeated dual (prediction) and primal (update) lifting steps. Hence, lifting is not only an alternative to classical filter banks but also offers the wavelet transform to work on non-dyadic and even irregular grids. In addition, lifting allows the reconstruction to be limited to integers as e. g. desired in digital image processing.

Donoho (1992) discusses several interpolating scaling functions, for instance of the Deslauriers-Dubuc wavelet basis or the spline wavelet basis (see also recent work by the Biomedical imaging group, Lausanne[7]). It holds that the autocorrelation function of an orthogonal scaling function is an interpolating scaling function. Even more, the autocorrelation of a Daubechies scaling function is a Deslauriers-Dubuc scaling function. Image interpolation is then achieved within the multiresolution analysis (MRA) framework.

[7] http://bigwww.epfl.ch/publications/

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Other authors (e. g. Su and Ward, 2006) use bi-orthogonal wavelets and rely on the smooth reconstruction filter for improved image quality. The given image is considered to be the low-resolution image, from which the detail components can be predicted in a suitable way. Adding the horizontal, vertical and diagonal edge information to the low-resolution image yields a high-resolution image of twice the size as the original.
All these interpolation approaches allow image enlargement up to a fixed size. Yet in view of the application of interest, a fiber tracking algorithm needs to be able to calculate the desired tensor information at any arbitrary point. These considerations have to be kept in mind for future work. In this thesis, I will not explore wavelet interpolation schemes but rather use, on top of voxelwise tensor estimation and wavelet based smoothing, quadratic interpolation as implemented in the available fiber tracking tool (Gössl et al., 2002).

4.2 Denoising the diffusion tensor

The discrete wavelet decomposition was reported to successfully denoise diffusion weighted images in combination with a Wiener filter (Wirestam et al., 2006). The approach follows the minority of smoothing methods for DTI data which operate right on the complex signal intensities before the Fourier transformation turns the Gaussian noise into a Rician distribution, in particular at locally low SNR (Gudbjartsson and Patz, 1995). Attempts have been made to smooth the magnitude images, too, with the focus on edge-preservation and statistical robustness (Hahn et al., 2001, details in Section 2.2). In this work, wavelet filtering is performed on processed data, namely on the array of diffusion tensor estimates. For this purpose, recall the estimation problem inherent to DTI. Formally, it is a matter of repeated measurement design

$$\mathbf{y}(s) \quad = \quad \mathbf{X}\boldsymbol{\beta}(s) + \boldsymbol{\varepsilon}(s), \quad \boldsymbol{\varepsilon}(s) \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \tag{4.10}$$

with $\boldsymbol{\beta}(s) = (\beta_1, \ldots, \beta_p)'(s)$ encoding for the $p = 6$ unknown elements that determine the symmetric covariance matrix at position $s, s = 1, \ldots, n$, in 3d space, i. e. the local diffusion tensor within the brain. Multiplication with the Moore-Penrose matrix inverse leads to the interpretation that the observed covariance entries $\mathbf{y}^*(s)$ are equal to the true

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

covariance entries $\boldsymbol{\beta}(s)$ corrupted by noise:

$$\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}(s)}_{\mathbf{y}^*(s)} = \boldsymbol{\beta}(s) + \boldsymbol{\varepsilon}^*(s), \quad \boldsymbol{\varepsilon}^*(s) \sim N(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \tag{4.11}$$

More precisely, the pretended observations $\mathbf{y}^*(s)$ are nothing else than the standard voxel-by-voxel estimates of the tensor elements, i. e. ST1 results. The aim is to recover for each tensor element the underlying ideal function $\beta_j(\boldsymbol{s}), \boldsymbol{s} \in \mathbb{R}^3, j = 1, \ldots, 6$, using DWT as implemented in the fast and memory-saving algorithm of Mallat.

Incorporation of the positive definiteness constraint

Log-Cholesky parametrization (Pinheiro and Bates, 1996) is one possibility to meet the positive-definiteness constraint of the diffusion tensors. Note that the restriction had not yet been imposed in the B-spline based SVCM, nor in ST1. Of course, immediate consideration through the estimation process would have been more appropriate. Cholesky factorization states that if and only if $\mathbf{R}$ is upper triangular with non-zero diagonal elements, then $\mathbf{R}'\mathbf{R}$ will be positive definite. Thus as long as the diagonal entries are exclusively negative or exclusively positive, the off-diagonal elements of the Cholesky factor $\mathbf{R}$ are unique. Either way, the intended uniqueness of $\mathbf{R}$ represents once more a constrained estimation problem. Yet if the diagonal shall comply with positive values, the logarithm can serve as auxiliary function to alter the parameter space unlimited on the real line. Hence, ensure positive definiteness of the diffusion tensor $\mathbf{D}$,

$$\mathbf{D} = \begin{pmatrix} \beta_1 & \beta_2 & \beta_4 \\ \beta_2 & \beta_3 & \beta_5 \\ \beta_4 & \beta_5 & \beta_6 \end{pmatrix} = \mathbf{R}'\mathbf{R} = \begin{pmatrix} r_1 & r_2 & r_4 \\ 0 & r_3 & r_5 \\ 0 & 0 & r_6 \end{pmatrix}' \begin{pmatrix} r_1 & r_2 & r_4 \\ 0 & r_3 & r_5 \\ 0 & 0 & r_6 \end{pmatrix}$$

by unique and simultaneously unconstrained log-Cholesky factorization:

$$\boldsymbol{\rho} = (\log(r_1), r_2, \log(r_3), r_4, r_5, \log(r_6))'.$$

Finally, reparameterize $\boldsymbol{\beta}$ as vector-valued function in $\boldsymbol{\rho}$ by definition of

$$\begin{aligned} \boldsymbol{\beta}(\boldsymbol{\rho}) = \big(&\exp(\rho_1)^2, \exp(\rho_1)\rho_2, \rho_2^2 + \exp(\rho_3)^2, \\ &\exp(\rho_1)\rho_4, \rho_2\rho_4 + \exp(\rho_3)\rho_5, \rho_4^2 + \rho_5^2 + \exp(\rho_6)^2\big)'. \end{aligned}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

In summary, the initial estimation problem (4.11) with $\boldsymbol{\beta}(s)$ being constrained to build a positive deifinite diffusion tensor can be rendered into an unconstrained estimation problem using log-Cholesky parametrization:

$$\boldsymbol{\rho}_{obs}(s) \;=\; \boldsymbol{\rho}_{true}(s) + \boldsymbol{\varepsilon}^{\sharp}(s), \quad \boldsymbol{\varepsilon}^{\sharp}(s) \sim N(\mathbf{0}, \mathrm{diag}(\sigma^2_{\sharp,j})). \tag{4.12}$$

The error term is an approximation in that independence is postulated for the transformed covariance elements $\rho_j, j = 1, \ldots, p$. With the log-Cholesky parametrization no further care is then needed in keeping track of the error term: The 3d wavelet decomposition can promptly be applied to each 3d field $\{\boldsymbol{\rho}_{obs,j}(s) : s = 1, \ldots, n\}$, where $\sigma^2_{\sharp,j}$ is estimated from the high-frequency weights of the first decomposition step. Hence, project elements $\boldsymbol{\rho}_{obs,j}, j = 1, \ldots, p$, onto 3d tensor product wavelets:

$$\boldsymbol{\rho}_{obs,j} \;=\; \mathbf{W}\boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}^{\sharp}_j.$$

The ordinary least squares estimate of the wavelet coefficient surfaces holds:

$$\boldsymbol{\gamma}^{OLS}_j \;=\; \underbrace{(\mathbf{W}'\mathbf{W})^{-1}}_{=\mathbf{I}}\mathbf{W}'\boldsymbol{\rho}_{obs,j}.$$

As explained in Section 4.1.4, the choice of a thresholding value and rule yields a sparse (and possibly shrunken) vector of wavelet contributions:

$$\hat{\boldsymbol{\gamma}}_{\lambda,j} \;=\; \delta_\lambda(\boldsymbol{\gamma}^{OLS}_j).$$

After having run the inverse discrete 3d wavelet transform, the regularized estimate

$$\hat{\boldsymbol{\rho}}_j \;=\; \mathbf{W}\hat{\boldsymbol{\gamma}}_{\lambda,j},$$

can be reparameterized to:

$$\hat{\boldsymbol{\beta}}_j \;=\; \boldsymbol{\beta}(\hat{\boldsymbol{\rho}}_j) = \begin{pmatrix} \exp(\hat{\rho}_1)^2 \\ \exp(\hat{\rho}_1)\hat{\rho}_2 \\ \hat{\rho}_2^2 + \exp(\hat{\rho}_3)^2 \\ \exp(\hat{\rho}_1)\hat{\rho}_4 \\ \hat{\rho}_2\hat{\rho}_4 + \exp(\hat{\rho}_3)\hat{\rho}_5 \\ \hat{\rho}_4^2 + \hat{\rho}_5^2 + \exp(\hat{\rho}_6)^2 \end{pmatrix}.$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Note that the competitive log matrix transformation (Pinheiro and Bates, 1996) is more stable but does not lead to the desired independence of the parametrized elements either. Certainly, the assumption of independently distributed errors in Eq. (4.12) is strong but indispensable for looping over the six 3d fields of parametrized tensor elements. In contrary, spatial correlation could be considered by special wavelet transforms for colored noise. Also level dependent thresholding is said to be more suitable in such a case.
In conclusion, the bunch of processing steps described above yields an adaptive estimator for smooth 3d components of positive definite covariance matrices.

4.3 Simulation study

4.3.1 Simulation model

The simulation model as presented in Fig. 3.5 is not apt for wavelet analysis. Of course, a mere resampling would satisfy the demand of dyadic dimensions. However, the miminum logarithm to the basis 2 of the 3d array size determines the maximum possible decomposition depth. For this reason, the spiral function is periodically extended along the z-axis prior to discrete sampling into $32 \times 32 \times 32$ voxels. As displayed in Fig. 4.13, the resulting imaging volume contains two concentrically nested spirals of five and twelve centimeters in radius, mimicking fiber bundles. Since each fiber-transit voxel is considered a fiber voxel, each transversal slice contains enough fiber structure (compare e. g. the projection on the xy-plane). The overall percentage amounts to 13.87% voxels. The degree of anisotropy in fiber as well as non-fiber voxels has been kept the same as in Section 3.4. Given the simulation model, $N = 100$ data sets were generated consisting of six diffusion weighted images corrupted by noise ($\sigma = 10$).

4.3.2 Study design

The parameter setup of the discrete wavelet analysis involves the wavelet family, the primary thresholding level j_0, the thresholding rule and the threshold itself. Boundary

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

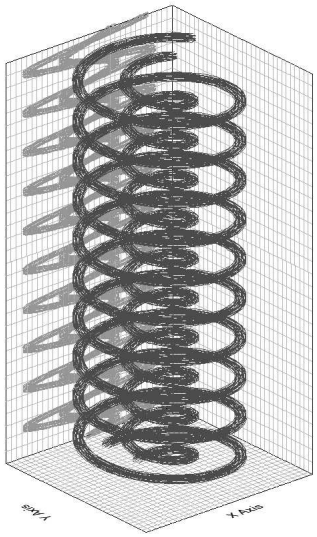


Figure 4.13: Geometrical model used for simulation studies with respect to wavelet analysis.

condition is not up for discussion as the employed R-package `waveslim` (Whitcher, 2005) handles border distortions only in a periodic fashion.

Since the basis function should mimic the underlying features of the given object, the choice of the wavelet family is crucial. Typical criteria are symmetry useful in avoiding dephasing in image processing, the number of vanishing moments important for compression purposes, and the regularity to get nice smoothness. Moreover, the length of the mother wavelet should not exceed the number of available observations. Apart from the Haar family, no wavelet system satisfies the demands of compact support (among others preferable for numerical stability) and symmetry at the same time. The latter condition is however indispensable in imaging if spatial directions are equally likely. Therefore, Daubechies *least asymmetric* wavelets with four vanishing moments (LA4) are investigated in addition to Haar (Haar) and Daubechies extremal phase wavelets (D2) with one and two vanishing moments, respectively. Also minimum bandwidth wavelets with four vanishing moments (MB4) are considered. The MB family is closely related to Daubechies extremal phase family, i. e. the members coincide for two vanishing moments and are almost the reverse of each other for more than two vanishing moments ($\phi_{Daub}(x) = -\phi_{MB}(x)$). The important difference consists in superior frequency-domain properties and, thus, superior edge-preserving property of MB wavelets compared to Daubechies wavelets given the same number of vanishing moments (Gençay et al., 2002, Chap. 4.3.3).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

The effect of the primary decomposition level was found to be neglibible according to a pilot study. Complete decomposition ($j_0 = 0$) is therefore performed.

Since hard thresholding can, in principle, be combined with the univeral or SURE threshold and since this seems more natural to a non-statistician (Donoho and Johnstone, 1994), it is taken into account. Thus, the comparison comprises four shrinkage methods: hard universal (hUNIV), soft universal (sUNIV), hard SURE (hSURE), and soft SURE (sSURE) thresholding. Every variant includes a noise estimate which, theoretically, can be level dependent and/or orientation dependent.

A preparatory study aims to investigate the impact of the noise estimation under the limitation to Haar wavelets. The estimation is either based on the highest resolution level (sigma1) or depends on each orientation (sigma2), each level (sigma3) or each level and each orientation (sigma4). As before, the quality is assessed via the error measures VMSE and AMSE (see Section 3.4, pp. 58, 60).
A second study concentrates on the impact of the wavelet type. For this purpose, the DWT of the simulated data is conducted using wavelets from the Haar (Haar) , the Daubechies extremal phase (D2), the least asymmetric (LA4), and the minimum bandwidth (MB4) family. The DWTs are then thresholded according to the above rules hUNIV, sUNIV, hSURE, and sSURE in combination with the two different noise estimates sigma1 and sigma2.

4.3.3 Simulation results

Concerning the preparatory study, Figures 4.14 and 4.15 display the results by means of boxplots. Clearly recognizable is the familiar distribution of smaller errors in background than in spiral voxels. The discrepancy between these two tissue classes becomes more pronounced with soft thresholding (second and fourth row from top). Concerning the noise estimation, level dependent approaches fail as expected from the simulation design that does not incorporate correlated noise. This is also true for the SURE approach, although whatsoever less evident when combined with the hard rule (third row from top). In contrast, the exclusive specification with respect to orientation (horizontal, vertical, diagonal details) shows hardly an effect compared to the uniform approach sigma1. A

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

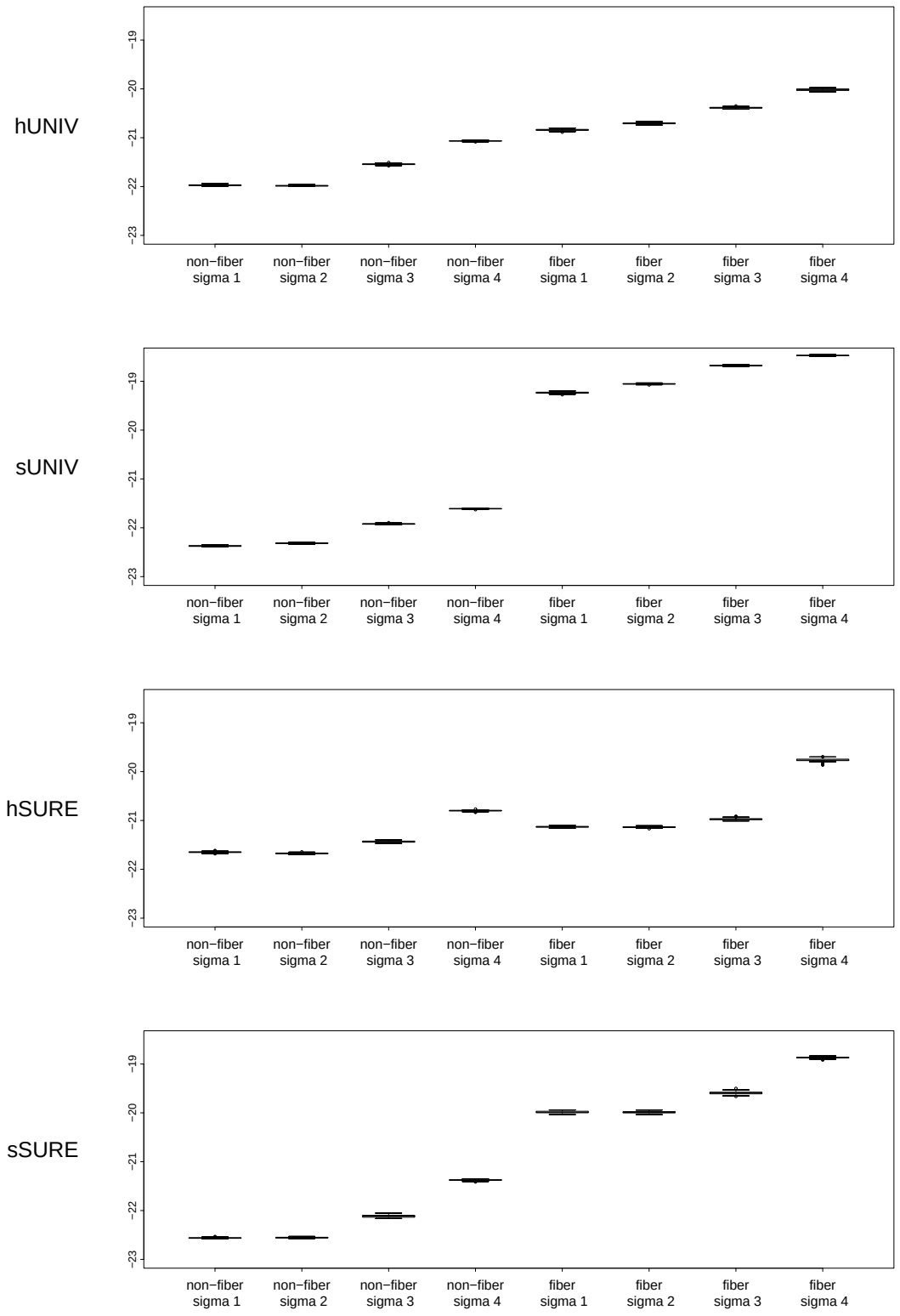


Figure 4.14: Boxplots of log AMSE for hUNIV, sUNIV, hSURE, and sSURE (top to bottom) with the noise estimated on the basis of the highest resolution level (sigma1), orientation dependent (sigma2), level dependent (sigma3), orientation and level dependent (sigma4) using Haar wavelets.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

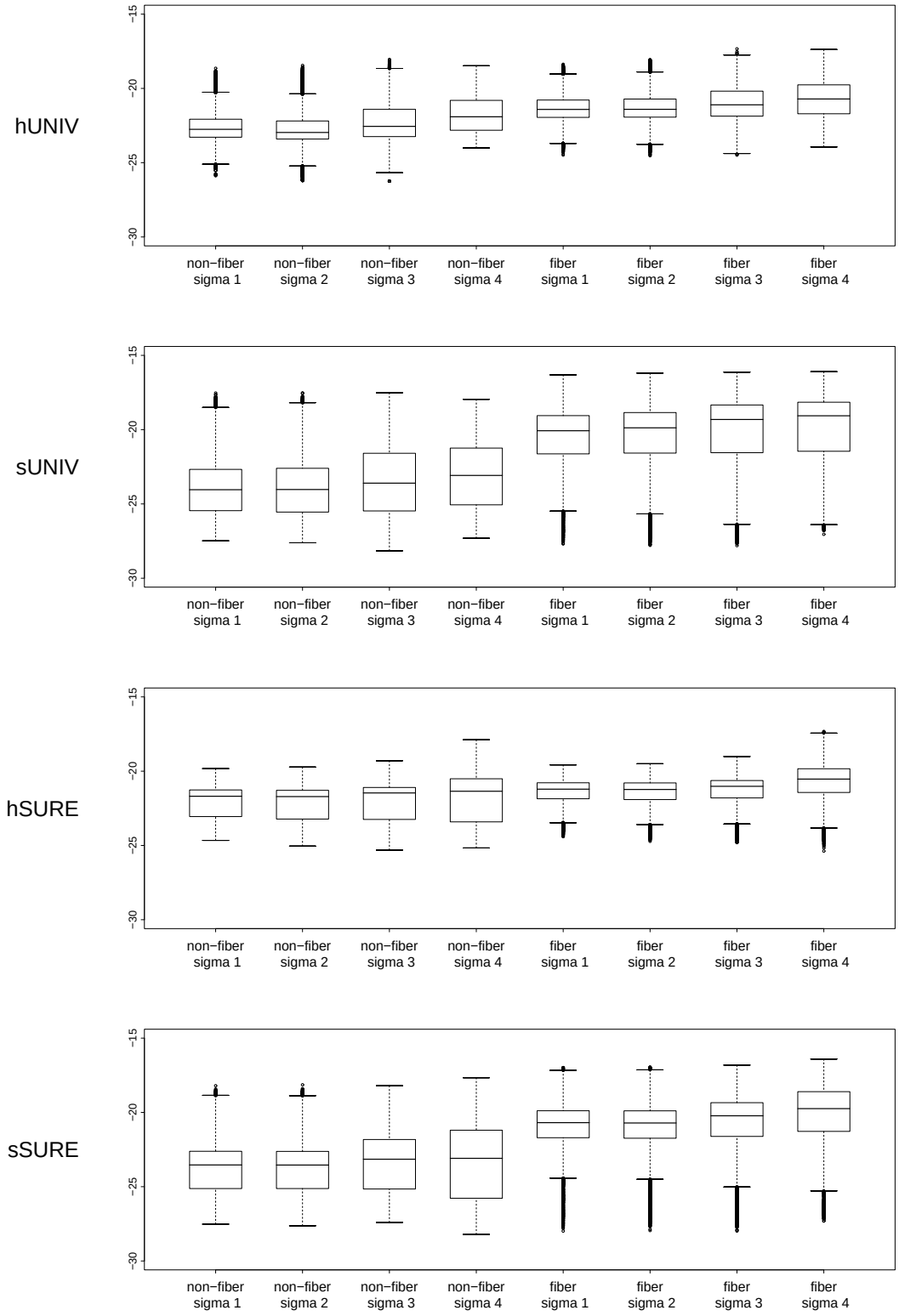


Figure 4.15: Boxplots of log VMSE for hUNIV, sUNIV, hSURE, and sSURE (top to bottom) with the noise estimated on the basis of the highest resolution level (sigma1), orientation dependent (sigma2), level dependent (sigma3), orientation and level dependent (sigma4) using Haar wavelets.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

slight improvement in terms of smaller log AMSE errors is visible for the SURE threshold applied either in a hard or soft way (Fig. 4.14, bottom rows).

Boxplots resulting from the second study design are not suited for visual inference because of apparent similarity. Tables 4.1 and 4.2 contain instead the principal boxplot statistics corresponding to the logarithmized average and voxelwise MSE, respectively.
At first glance, the wavelet type plays hardly a role, in particular with respect to the errors of D2, LA4 and MB4. The resembling behavior points at an oversized filter length given the scale of the features to be preserved. In view of this, especially the minimum bandwidth wavelet is not able to develop its superior edge-preserving property. The Haar wavelet has the shortest support among the examined families and can therefore still capture substantial structures at finer scales. Not surprisingly, the Haar wavelet performs best under almost all shrinkage conditions in terms of an overall error measurement (log AMSE). Regarding the local error (log VMSE), the D2 family outperforms the Haar family in background voxels but not in spiral voxels.
The performance on these two tissue types is quite similar if hard thresholding is applied. In contrast, soft thresholding leads to distinctively larger errors in fiber associated regions. At the same time, the shrinkage entails a smooth and less erroneously reconstructed background. This holds for all parameter settings and both error measures.
Log AMSE errors obtained with the SURE threshold tend to be smaller than those of universal thresholding. If the focus is on log VMSE errors, universal hard thresholding seems to be preferrable over the SURE counterpart.
The choice of the noise estimate does, in general, not affect the MSE ranking of the wavelet types (compare black and gray prints). This result corroborates the previous finding obtained from analyses with the Haar wavelet.

4.3.4 Comparison to standard approaches

Coming back to the goal of improving regularized tensor estimation by modified DTI data processing, the comparison to previous methods is essential. Figure 4.16 therefore depicts the error boxplots of voxelwise regression without (ST1) and with subsequent Gaussian smoothing (ST2), sequential SVCM with linear B-splines, one knot per 1.28 voxels and

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
hUNIV	maximum	-21.99	-21.69	-21.65	-21.55	-20.89	-20.60	-20.61	-20.59
		-22.00	-21.62	-21.59	-21.49	-20.74	-20.37	-20.38	-20.34
	upper hinge	-21.98	-21.67	-21.64	-21.54	-20.85	-20.58	-20.59	-20.56
		-21.99	-21.62	-21.59	-21.47	-20.71	-20.35	-20.36	-20.31
	median	-21.97	-21.67	-21.64	-21.53	-20.84	-20.56	-20.58	-20.55
		-21.98	-21.61	-21.58	-21.47	-20.71	-20.34	-20.35	-20.30
	lower hinge	-21.96	-21.66	-21.63	-21.53	-20.83	-20.55	-20.57	-20.53
		-21.98	-21.61	-21.58	-21.46	-20.70	-20.33	-20.34	-20.29
	minimum	-21.94	-21.65	-21.62	-21.51	-20.80	-20.52	-20.54	-20.49
		-21.96	-21.59	-21.57	-21.45	-20.67	-20.31	-20.32	-20.27
		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
sUNIV	maximum	-22.39	-22.07	-22.00	-21.88	-19.28	-19.14	-19.20	-19.19
		-22.33	-21.96	-21.90	-21.80	-19.08	-18.86	-18.90	-18.87
	upper hinge	-22.38	-22.05	-21.99	-21.87	-19.24	-19.11	-19.17	-19.16
		-22.32	-21.95	-21.88	-21.79	-19.06	-18.84	-18.88	-18.86
	median	-22.37	-22.04	-21.98	-21.86	-19.23	-19.10	-19.16	-19.15
		-22.32	-21.94	-21.88	-21.78	-19.05	-18.84	-18.88	-18.85
	lower hinge	-22.36	-22.03	-21.97	-21.86	-19.22	-19.09	-19.15	-19.14
		-22.31	-21.94	-21.87	-21.78	-19.05	-18.83	-18.87	-18.85
	minimum	-22.35	-22.02	-21.96	-21.85	-19.20	-19.07	-19.12	-19.12
		-22.29	-21.93	-21.86	-21.76	-19.03	-18.82	-18.86	-18.83
		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
hSURE	maximum	-21.68	-21.59	-21.58	-21.53	-21.16	-21.02	-21.03	-21.02
		-21.70	-21.62	-21.61	-21.56	-21.17	-21.00	-21.01	-20.99
	upper hinge	-21.66	-21.57	-21.56	-21.51	-21.14	-20.99	-21.00	-20.99
		-21.68	-21.60	-21.59	-21.54	-21.14	-20.98	-20.99	-20.96
	median	-21.65	-21.56	-21.55	-21.51	-21.13	-20.99	-21.00	-20.98
		-21.67	-21.60	-21.58	-21.54	-21.14	-20.97	-20.98	-20.95
	lower hinge	-21.64	-21.56	-21.54	-21.50	-21.12	-20.98	-20.99	-20.97
		-21.67	-21.59	-21.58	-21.53	-21.13	-20.96	-20.97	-20.93
	minimum	-21.62	-21.53	-21.52	-21.48	-21.10	-20.95	-20.96	-20.91
		-21.64	-21.57	-21.56	-21.51	-21.11	-20.93	-20.94	-20.90
		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
sSURE	maximum	-22.58	-22.28	-22.22	-22.11	-20.04	-19.82	-19.87	-19.80
		-22.57	-22.27	-22.21	-22.09	-20.04	-19.78	-19.82	-19.74
	upper hinge	-22.57	-22.27	-22.21	-22.09	-20.00	-19.76	-19.79	-19.75
		-22.56	-22.25	-22.19	-22.08	-20.00	-19.75	-19.76	-19.72
	median	-22.56	-22.26	-22.20	-22.09	-19.99	-19.75	-19.78	-19.73
		-22.55	-22.25	-22.19	-22.07	-19.99	-19.73	-19.75	-19.70
	lower hinge	-22.56	-22.25	-22.20	-22.08	-19.97	-19.73	-19.76	-19.71
		-22.55	-22.24	-22.18	-22.06	-19.97	-19.72	-19.73	-19.68
	minimum	-22.53	-22.24	-22.19	-22.06	-19.94	-19.70	-19.73	-19.63
		-22.53	-22.23	-22.18	-22.04	-19.94	-19.68	-19.69	-19.65

Table 4.1: Boxplot statistics of log AMSE for different thresholding rules with either sigma1 (black) or sigma2 (gray) noise estimation using different wavelet families.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
hUNIV	maximum	-25.87	-26.97	-26.39	-25.97	-24.49	-24.42	-24.37	-24.22
		-26.20	-27.29	-26.76	-26.14	-24.53	-24.59	-24.39	-24.31
	upper hinge	-23.28	-23.23	-23.19	-23.14	-21.95	-21.86	-21.93	-22.00
		-23.41	-23.37	-23.32	-23.25	-21.93	-21.80	-21.86	-21.93
	median	-22.75	-22.57	-22.50	-22.33	-21.41	-20.93	-20.98	-21.00
		-22.97	-22.76	-22.67	-22.48	-21.41	-20.83	-20.86	-20.90
	lower hinge	-22.07	-21.46	-21.38	-21.20	-20.78	-20.31	-20.35	-20.36
		-22.19	-21.47	-21.39	-21.19	-20.72	-20.11	-20.15	-20.15
	minimum	-18.64	-18.73	-18.45	-18.26	-18.38	-17.90	-17.68	-17.50
		-18.45	-18.28	-18.40	-18.03	-18.06	-17.57	-17.37	-17.10
		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
sUNIV	maximum	-27.48	-28.49	-28.29	-27.56	-27.68	-27.61	-27.63	-27.38
		-27.62	-28.59	-28.45	-27.70	-27.79	-27.72	-27.79	-27.43
	upper hinge	-25.45	-25.64	-25.48	-25.35	-21.63	-21.57	-21.66	-21.71
		-25.55	-25.74	-25.56	-25.44	-21.58	-21.49	-21.58	-21.60
	median	-24.04	-24.03	-23.87	-23.65	-20.07	-20.06	-20.17	-20.23
		-24.03	-24.05	-23.91	-23.71	-19.88	-19.76	-19.86	-19.84
	lower hinge	-22.67	-22.24	-22.25	-22.06	-19.06	-18.83	-18.88	-18.91
		-22.60	-22.17	-22.18	-21.99	-18.85	-18.53	-18.56	-18.56
	minimum	-17.53	-17.74	-17.73	-17.27	-16.31	-16.39	-16.49	-16.23
		-17.53	-17.63	-17.80	-17.23	-16.19	-16.18	-16.31	-16.11
		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
hSURE	maximum	-24.66	-25.04	-24.53	-24.43	-24.41	-24.86	-24.61	-24.31
		-25.05	-25.14	-24.94	-24.72	-24.72	-25.30	-24.92	-24.57
	upper hinge	-23.06	-23.07	-23.07	-23.06	-21.86	-21.84	-21.89	-21.97
		-23.22	-23.21	-23.25	-23.26	-21.91	-21.93	-22.00	-22.11
	median	-21.69	-21.69	-21.69	-21.68	-21.21	-21.09	-21.09	-21.09
		-21.71	-21.76	-21.75	-21.75	-21.23	-21.11	-21.11	-21.10
	lower hinge	-21.26	-21.19	-21.15	-21.05	-20.78	-20.61	-20.63	-20.63
		-21.29	-21.23	-21.20	-21.10	-20.79	-20.60	-20.62	-20.62
	minimum	-19.82	-19.78	-19.59	-19.43	-19.57	-19.19	-18.99	-18.74
		-19.73	-19.54	-19.44	-19.18	-19.50	-18.91	-18.81	-18.55
		non-fiber				fiber			
		Haar	D2	LA4	MB4	Haar	D2	LA4	MB4
sSURE	maximum	-27.52	-28.42	-28.05	-27.84	-27.99	-27.84	-28.00	-27.64
		-27.63	-28.33	-28.14	-28.02	-27.95	-27.95	-27.75	-27.74
	upper hinge	-25.13	-25.35	-25.26	-25.14	-21.70	-21.61	-21.69	-21.76
		-25.13	-25.31	-25.25	-25.11	-21.74	-21.66	-21.74	-21.82
	median	-23.54	-23.55	-23.46	-23.30	-20.69	-20.53	-20.63	-20.66
		-23.54	-23.60	-23.51	-23.33	-20.71	-20.54	-20.63	-20.67
	lower hinge	-22.62	-22.20	-22.17	-22.01	-19.89	-19.58	-19.60	-19.57
		-22.62	-22.21	-22.19	-22.02	-19.89	-19.56	-19.57	-19.54
	minimum	-18.20	-18.30	-18.19	-17.74	-16.98	-16.92	-16.96	-16.68
		-18.14	-18.21	-18.08	-17.66	-16.95	-16.85	-16.93	-16.59

Table 4.2: Boxplot statistics of log VMSE for different thresholding rules with either sigma1 (black) or sigma2 (gray) noise estimation using different wavelet families.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

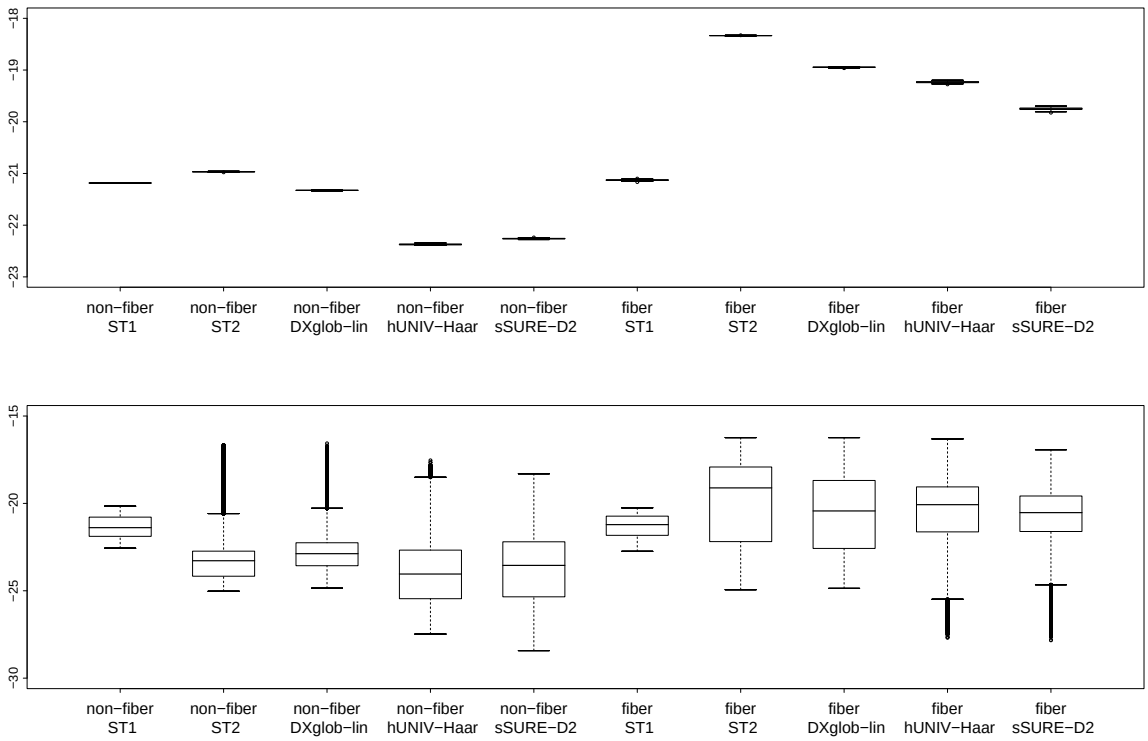


Figure 4.16: Boxplots of log AMSE (top) and log VMSE (bottom) for ST1, ST2, DXglob-lin, hUNIV-Haar and sSURE-D2.

global tuning parameter for first order difference penalties (DXglob-lin), as well as ST1 filtered once based on Haar wavelets combined with universal thresholding (hUNIV-Haar) and once based on Daubechies wavelets combined with SURE shrinkage (sSURE-D2). Noise estimation conforms to sigma1.

Clearly, both wavelet transforms convince with satisfactory error reduction. The improvement appears larger from an overall (top row) than from a local perspective (bottom row). In particular, errors in fiber voxels can be lowered using SURE shrinkage, outperforming ST2 and the B-splines based SVCM, but not ST1. The voxelwise least squares estimates show approximately the same error distribution regardless the underlying degree of anisotropy. This can be considered as proof of concept, since globally superimposed noise prolongates to the same amount into local estimates.

The VMSE errors, collapsed into distributional characteristics in Fig. 4.16, can alternatively be presented as maps to enable local assignment. For this purpose, log ratios of $\text{VMSE}_{\text{method 1}}$ and $\text{VMSE}_{\text{method 2}}$ are build as in Section 3.4. Where previously results from the B-spline based SVCM were contrasted with those obtained by ST2

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

(Fig. 3.8), wavelet filtered tensor estimates are now related to ST2 and to each other. Figure 4.17 contains the maps of $\log(\mathrm{VMSE}_{\text{hUNIV-Haar}}/\mathrm{VMSE}_{\text{sSURE-D2}})$ in the upper right, $\log(\mathrm{VMSE}_{\text{hUNIV-Haar}}/\mathrm{VMSE}_{\text{ST2}})$ in the lower left and $\log(\mathrm{VMSE}_{\text{sSURE-D2}}/\mathrm{VMSE}_{\text{ST2}})$ in the lower right corner. The interpretation is as follows: Rose colors indicate superiority of the numerator, while green colors hint to a better performance of denominator method. As obvious from the rose color dominance, both wavelet approaches outperform the standard voxelwise regression followed by Gaussian smoothing (bottom subfigures). In the interspiral space, however, ST2 shows higher accuracy. This becomes particularly evident

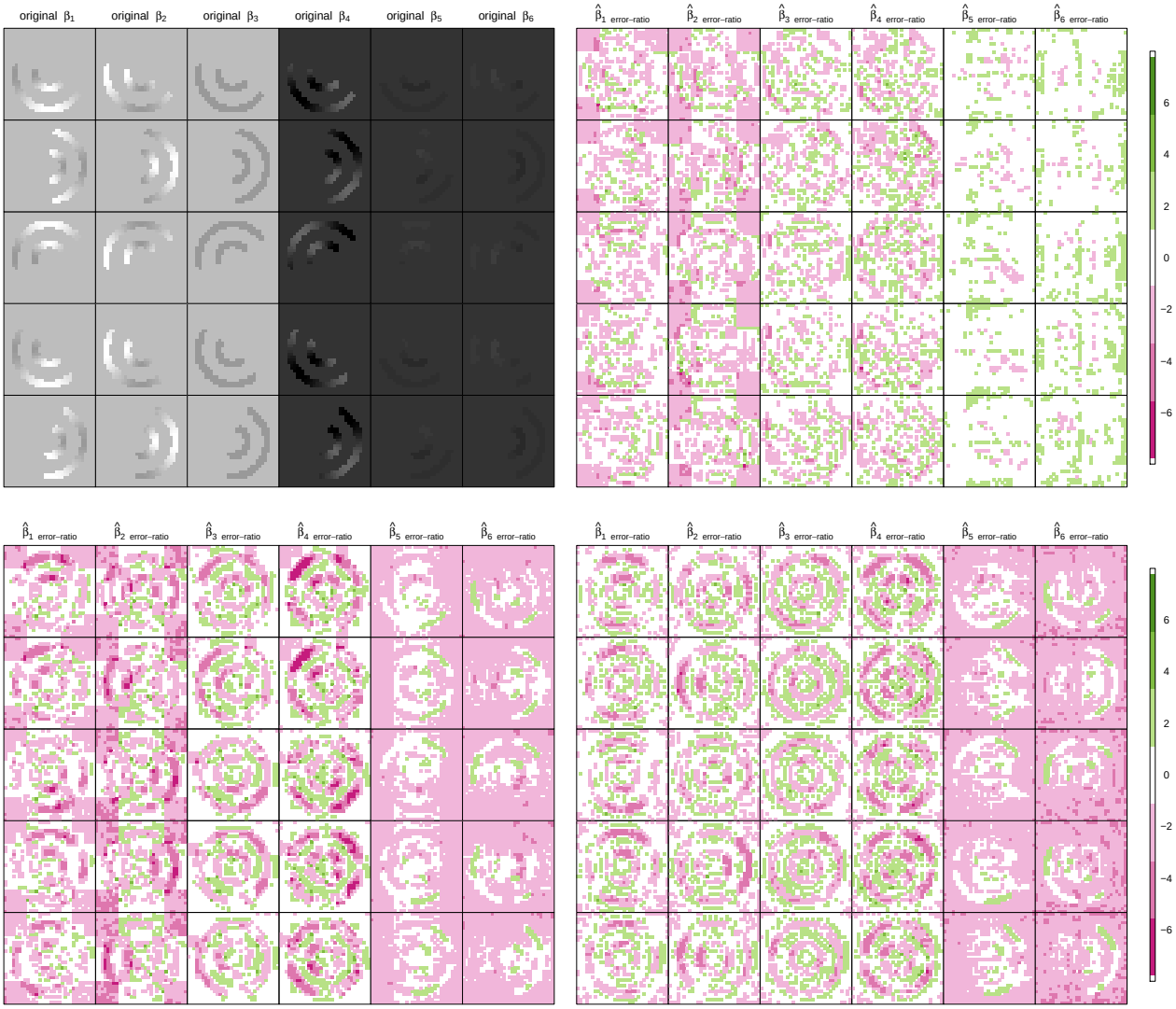


Figure 4.17: Template coefficient surfaces (top left) correspond to the six elements of the 3d tensor field. Log ratio of VMSE is given for both hUNIV-Haar (bottom left) and sSURE-D2 approach (bottom right) relative to ST2. Top right shows the log ratio of VMSE from hUNIV-Haar versus sSURE-D2. All subfigures show exemplary middle slices 14 to 18.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

when compared to sSURE-D2 (bottom right). The green pattern furthermore gives rise to the presumption of Gibb's artifacts caused by sSURE-D2. This phenomenon might be traceable in the estimates themselves (Fig. 4.18). In general, the wavelet filters seem to outperform the B-spline SVCMs with respect to background smoothness. This can be concluded without explicit comparison of the two methods (B-splines or wavelets), since ST2 represents the benchmark in both Fig. 3.8 and Fig. 4.17. However, the behavior at ridges is not as stable as in the B-spline case. For example, the wavelet approaches have difficulties to reproduce signals of low intensity (green spots in the fifth and sixth column of bottom maps).
As it can be judged from the upper right subfigure, hUNIV-Haar performs better at edges and yields smoother background in the first and second diagonal element, whereas sSURE-D2 primarily succeeds better in catching up the signal magnitude at ridges.
The white cross-shaped pattern hints to the presence of *blocking artifacts*, i. e. effects at different scales that are due to the squared support of the Haar wavelet. Blocking artifacts do not emerge with the D2 wavelet (bottom right).

Facing ambigously informative figures of the error ratios, a look onto the estimated diffusion tensors themselves seems worthwhile. For display feasibility, the focus is limited on the middle slices 14 to 18 and on the first diagonal tensor element for which error extremes as well as Gibbs and blocking distortions are both present (see Fig. 4.17). Figure 4.18 compares the original noise-free template with the standard approaches, the SVCM method and the wavelet filters applied to data from one exemplary simulation run.

Depending on the printing quality, all method specific drawbacks and advantages are more or less apparent at a glance:

ST1 Mere voxelwise regression completely lacks spatial smoothness though the signal magnitude seems to be caught satisfactorily in fiber voxels.

ST2 Consecutive application of a Gaussian kernel with FWHM equal to 0.75 voxels seriously oversmoothes, in particular in z-direction.

DXglob-lin Space-varying coefficient surfaces designed with penalized B-splines are less overregulated, contrasts are enhanced compared to ST2. Gibbs-like shadows are reduced to a minimum using linear basis functions.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

hUNIV-Haar Features are correctly mapped regarding brightness and edge location. Smoothing is manifest but at the expense of blocking artifacts. For instance, partial failure of the spiral pathway happens at eight o'clock, top row. In the background, the Haar basis function is even fully recognizable (compare to Fig. 4.8). Gibbs phenomena are best visible as black voxels neighboring the bright curve segments.

sSURE-D2 For the first time, edge preservation appears satisfactorily balanced to smoothing. The interspiral space is slightly more turbulent for the D2 than for the Haar approach. Gibbs artifacts are still present, whereas blocking does hardly disturb the visual impression.

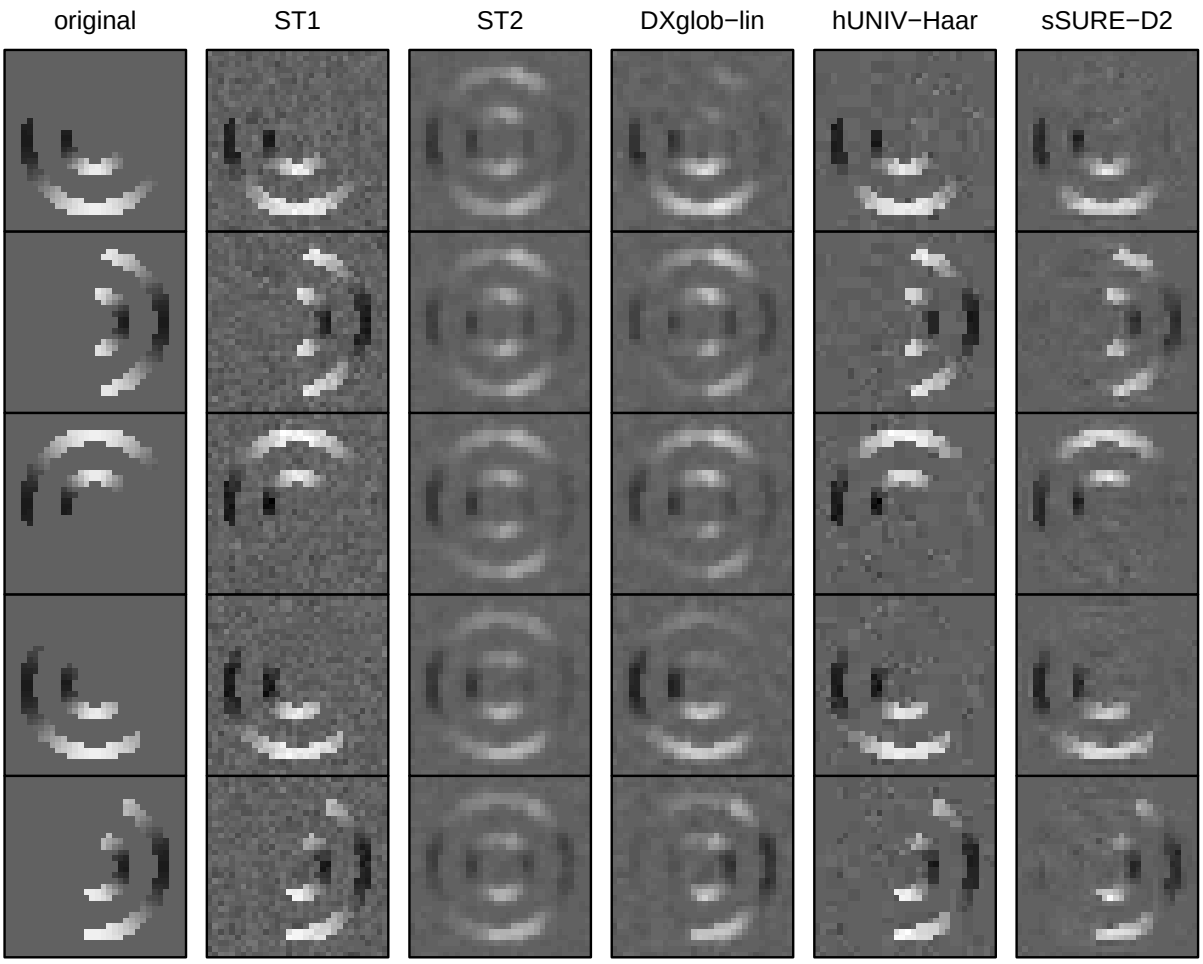


Figure 4.18: Surface excerpt (slices 14 to 18) showing the first diagonal diffusion tensor element as noise-free template (left) and estimated using voxelwise regression (ST1) with subsequent Gaussian smoothing (ST2) and space-varying coefficient approaches based on B-splines (DXglob-lin), Haar (hUNIV-Haar) or Daubechies wavelets (sSURE-D2).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

4.4 Real life example

In this section, an exemplary DTI data set of the human brain is processed using the wavelet filter that is rated best in the above simulations. For this purpose, the input data consists of six measurement repeats, each comprising 16 successive slices of originally 128×128 voxels. Although in-plane decomposition can in principle be determined by $j_0 = 7$, the possible maximum decomposition level in z-direction limits the thresholding to coefficients at the first four levels. This way, a realistic noise cut-off frequency is met at least within plane. Given the structural variety and the larger dimensions of real data, symmlets and minimum bandwidth wavelets serve for re-examination. The analysis is therefore conducted with the wavelet families D2, LA4 as well as MB4 together with SURE shrinkage and level independent noise estimate (sigma1). After tensor estimation, the volume is resized to the same six slices of 90×75 voxels as used in Chapter 3.5.
A main difficulty, that did not occur with simulated samples, results from log-Cholesky parametrization of the voxel-by-voxel estimated tensor data (Eq. (4.11)). In few voxels, the Cholesky decomposition is not feasible straight away. Owing to the various sources of noise in MR imaging or to signal drop, the ST1 based covariance matrices are singular in about 0.5% of all 40500 voxels. For the locations of the affected voxels refer to Fig. 4.19.

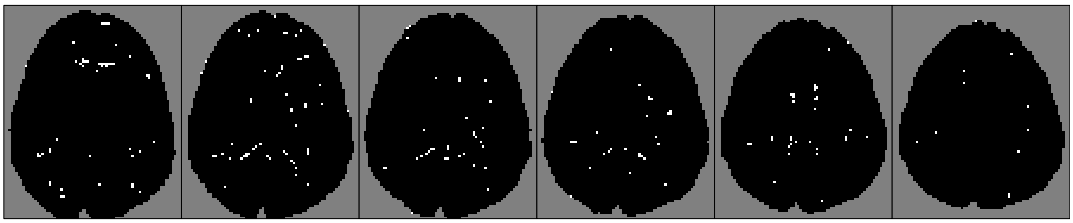


Figure 4.19: Mask of six middle human brain slices with problematic voxels superimposed in white. These indicate where the voxelwise estimates of the diffusion tensor are singular.

According to Skare et al. (2000b), negative eigenvalues are more likely with increasing noise and increasing FA. The latter statement holds for most of the affected voxels that seem to lie at the boundary to the ventricles where highly organized white matter is in direct neighborhood (Fig. 4.19). Hence, ensuring positive definiteness is particularly important with respect to fiber tracking. As remedy in such cases, small positive amounts are added on the diagonal elements prior to Cholesky decomposition.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

The problem of visualizing an estimated tensor field is equivalent to the problem of visualizing a vector field along with a scalar function on the same plot. Apart from showing the principal eigenvectors together with a color-code for the degree of local anisotropy, FA color-encoded ellipsoids are most intuitive and also somewhat more informative at first sight. For example, planar-shaped tensors may hint to an underlying fiber crossing or branching. Figure 4.20 was generated using the software module 'TensorViewer' developed by (Fillard and Toussaint, 2006).

Clearly, ST1 leads to the most speckled impression (Fig. 4.20 (a)). Interestlingly, the result is hardly discernible from the filtered version involving D4 wavelets (Fig. 4.20 (d)). The

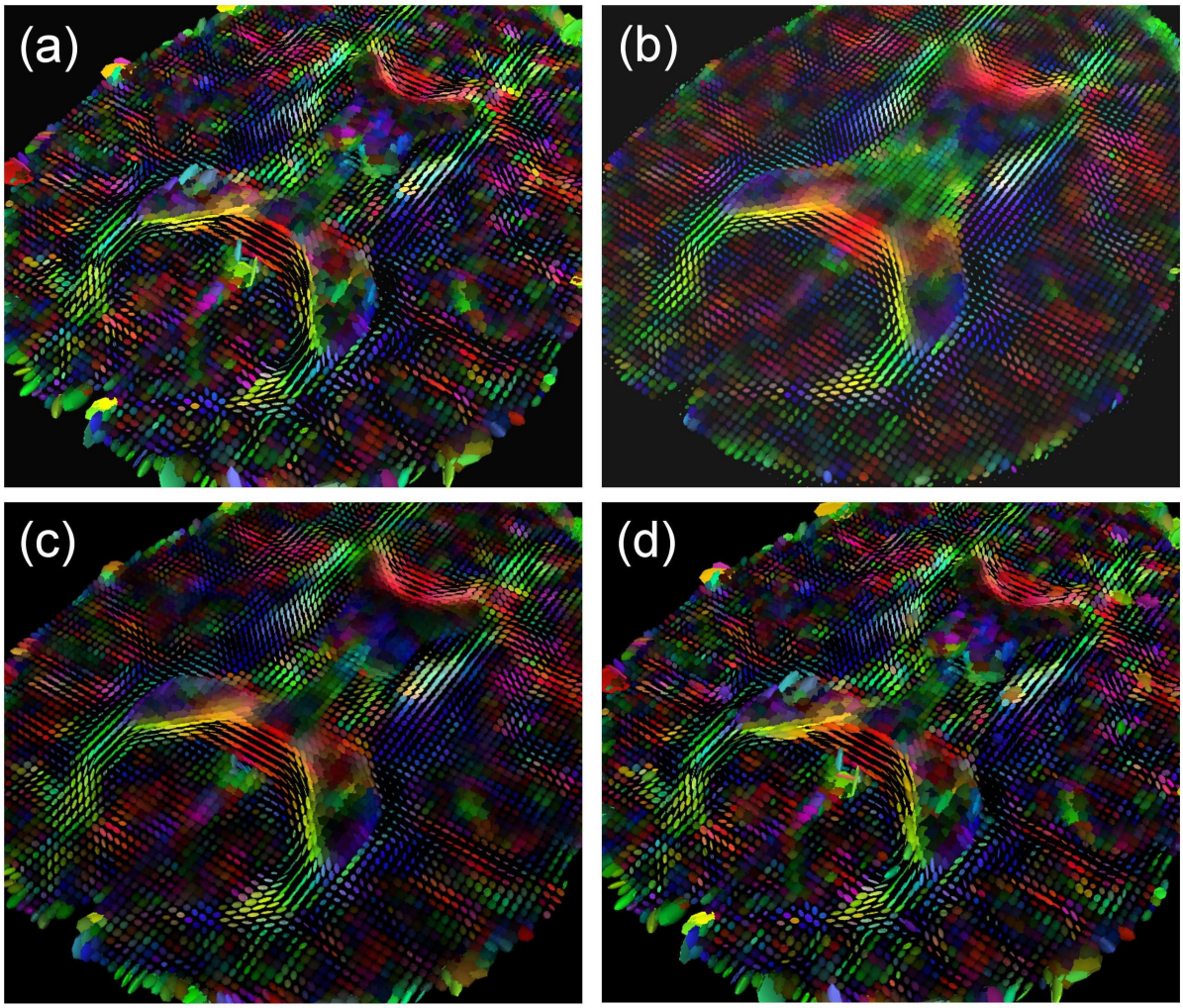


Figure 4.20: An exemplary slice of the 3d diffusion tensor field estimated from human DTI data using (a) ST1, (b) ST2, (c) DXglob-lin and (d) sSURE-D4, and displayed as FA color-encoded ellipsoids. The colorization is related to the principal diffusion direction and, thus, indicates left-right (red), anterior-posterior (green), or superior-inferior (blue) orientation of the associated fiber bundle. Intermediate directions are represented by the corresponding combination color.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

disturbing effect is triggered by a few ellipsoids of bigger size than average, in particular present at the slice center and border. These tensors can be assigned to full or partial volumes of cerebro-spinal fluid (CSF), e. g. to the central ventricles where diffusion is free. Usually, a mask, built on the basis of the b_0 image, is applied to sort out CSF. In addition, these voxels are often excluded from fiber tractography. Since no masking was performed prior to estimation, the wavelet approach smears the ventricle compartement into the splenium of the corpus callosum, i. e. into the green-yellow-red-yellow-green c-shape. This result is contrary to what is anticipated from wavelet theory and lacks justification. Also, regularization is hardly given though presumably paramount for fiber tractography. The results obtained by LA4 and MB4 wavelets do not differ significantly. In contrast, the tensor fields in Fig. 4.20 (b) and (c) convince by apparent smoothness. However, Gaussian kernel smoothing (Fig. 4.20 (b)) leads to overblurring in comparison to the SVCM with penalized B-splines. For example, the red-colored voxels in genu and splenium of the corpus callosum are less cigar-shaped though this region is highly anisotropic. Also in red, u-shaped region in grey matter contain more planar ellipsoids.

4.5 Discussion

Various simulation studies have been conducted accounting for different noise estimates, the use of different wavelet families and the application of different thresholding values and thresholding rules. As expected, scale dependent noise estimation proved to be inadequate to simulated data imposed with Gaussian white noise. Hard thresholding allows to reproduce structural aspects in almost full accuracy while uniform regions are still subjected to some error. Soft thresholding generally leads to smoother results, i. e. a correctly smooth background but also slightly blurred features. Depending on the application of interest, the soft shrinkage results are clearly to be favored with their moderate and adaptive regularization. The use of Haar wavelets should be avoided since undesired blocking effects are likely to emerge and the outcome equals a (higher-dimensional) step function. Least asymmetric separable 3d wavelets as well as minimum bandwidth wavelets showed neither beneficial nor detrimental over the use of Daubechies extremal phase family.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

In contrast to former methods such as simple kernel smoothing or the B-spline based SVCM, the multiscale analysis with wavelets yields significantly less erroneous results due to the wavelets' built-in spatial-adaptivity. Space-frequency resolution particularly matters for the sensitivity to local curvature changes. B-splines (degree > 0), for example, cannot model aprupt changes due to their wide and overlapping support regions.
However, comparative error maps pointed to Gibbs-like phenomena accompanying wavelet filtering. Although these Gibbs artifacts are much more damped compared to the output from B-spline SVCMs, orthogonal wavelet analysis is not able to remedy them completely. Jansen (2001, p. 120) derives a theoretical explanation for the oscillation causes by looking at the optimal and experimental thresholding rules. According to (Jansen, 2001), hard thresholding is more biased than soft thresholding and, thus, more likely to develop over- and undershoots near singularities. Yet, in the given simulation study, also the wavelet reconstruction using SURE shrinkage shows oscillatory behavior at fiber rims. Another reasoning considers the lack of translation invariance inherent to the DWT as origin of Gibbs effects (see Ogden (1997, p. 183) and the references therein). The use of the translation invariant wavelet transform (MODWT) may therefore be recommended.

Rather bizarre appears the fact that an SVCM involving Haar wavelets yields the best reconstruction in terms of MSE. The discrepancy between the resulting 3d step function and the intended smooth surface tackles the basic problem of the most relevant measure of error. Obviously, the "mean of squared errors (MSE) does not correspond to a human perception of generality. Since our visual system seems to work on a multiscale basis, a norm based on a multiresolution decomposition might be a better expression of visual quality" (Jansen, 2001, p. 49). In conclusion, the problem raises the question after the usefulness of theoretical optimality for experimental setups. Several proposals to encounter this "paradoxon" are made in Marron and Tsybakov (1995).

Finally, note that smaller errors of the tensor estimates do not necessarily imply smaller errors of the principal eigenvectors. Due to the non-linearity of the eigen decomposition, directional information of the main diffusivity could remain stable despite large AMSE and VMSE errors of the tensor entries. Recall that any tracking algorithm primarily exploit the eigenvectors. Hence, future studies should at least investigate the angular distribution between noise-free and reconstructed eigenvectors as well.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Concerning the human brain example, the SVCM on the basis of penalized linear B-splines yields the best visual result. Gaussian kernel smoothing leads to oversmoothing, as excepted, and voxelwise tensor estimation is obviously not apt for the ultimate goal of fiber tracking. Against the theoretical background and the results obtained in the simulation study, the wavelet approach fails in producing an appropriately regularized tensor field. Moreover, the edge-preserving property is challenged by unmasked CSF voxels. This disappointing result calls for further investigation of available wavelet transforms and thresholding mechanisms. Last but not least, the wavelet procedure needs to be adapted to the non-unit voxel size that is common in practice. Either the use of an anisotropic basis or a scaling step seems advisable.

4.6 Proposal for a wavelet based SVCM

Ordinary least squares regression represents the standard approach to assess the diffusion tensor voxel-by-voxel. Subsequent wavelet filtering proved capable to preserve edges, while a low smoothing effect could be detected from the real life example (Fig. 4.4). More experience with this kind of basis functions is therefore needed, in particular in view of tractography. For the time being, let me sketch a possible solution to the original problem. The goal is to improve the B-spline based SVCM by substituting B-splines with wavelets. It has been substantially argued in Section 3.1 that the covariance matrix elements in Eq. (3.5) are preferentially modelled non-parametrically as space-varying coefficient surfaces:

$$\boldsymbol{\beta}_j \quad = \quad \mathbf{B}\boldsymbol{\gamma}_j, \quad j = 1, \ldots, p.$$

The matrix $\mathbf{B}$ stands for B-splines and contains the basis functions evaluated at grid points $s, s = 1, \ldots, n$, as employed previously. A naive proposal would simply exchange $\mathbf{B}$ with the matrix $\mathbf{W}$ of wavelet basis functions. This is however hardly recommendable with Mallat's efficient pyramidal algorithm at hand (p. 98). Using the filterbank implementation, merely the wavelet coefficients are computed but not the wavelets themselves. Also, an analytical expression of the basis functions is often not available. Hence, the proposal must "eliminate" the matrix $\mathbf{W}$ as pointed out below.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

A rearrangement of the r repeatedly measured $(n_1 \times n_2 \times n_3)$-dimensional diffusion weighted images into the vector $\mathbf{y}$,

$$\mathbf{y} = (\mathbf{y}_1{}', \ldots, \mathbf{y}_r{}')' = (y_{1,1}, \ldots, y_{n,1}, \ldots\ldots, y_{1,r}, \ldots, y_{n,r})', \quad n = n_1 n_2 n_3,$$

causes a swap of the Kronecker product in the joint model (Eq. (3.5)) for the n separate regression equations of the standard approach (ST1):

$$\mathbf{y} \;=\; \sum_{j=1}^{p} \big(\mathbf{X}(\cdot, j) \otimes \mathbf{I}_n\big)\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_{rn}). \tag{4.13}$$

The jth vectorized surface $\boldsymbol{\beta}_j$, $\boldsymbol{\beta}_j = (\beta_j(1), \ldots, \beta_j(n))'$, is projected onto 3d tensor product wavelets:

$$\boldsymbol{\beta}_j \;=\; \mathbf{W}\boldsymbol{\gamma}_j, \quad j = 1, \ldots, p,$$

where $\mathbf{W}$ is $(n \times n)$-dimensional and encodes for the matrix of evaluated wavelet basis functions. In an image of dyadic dimensions, there are always as many wavelets as grid points. This allows to restate Eq. (4.13):

$$\mathbf{y} \;=\; (\mathbf{X} \otimes \mathbf{W})\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{4.14}$$

with $(r \times p)$-dimensional regressor matrix $\mathbf{X}$ and $(pn \times 1)$-dimensional vector of wavelet coefficients $\boldsymbol{\gamma}$. For further reformulation, recall the basic properties of the Kronecker product (Dierckx, 1993, p. 170), where the vector operator $\mathrm{vec}(\mathbf{X})$ returns the columnwise stacked matrix $\mathbf{X}$:

$$(\mathbf{A} \otimes \mathbf{B})' \;=\; \mathbf{A}' \otimes \mathbf{B}' \tag{4.15}$$

$$(\mathbf{A} \otimes \mathbf{B})^{-1} \;=\; \mathbf{A}^{-1} \otimes \mathbf{B}^{-1} \tag{4.16}$$

$$(\mathbf{AB}) \otimes (\mathbf{CD}) \;=\; (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D}) \tag{4.17}$$

$$\mathrm{vec}(\mathbf{X}_{n\times m}\, \mathbf{B}_{m\times q}) \;=\; (\mathbf{B}' \otimes \mathbf{I}_n)\, \mathrm{vec}(\mathbf{X}) \tag{4.18}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Subsequently, the OLS estimate of $\boldsymbol{\gamma}$ is derived from Eq. (4.14):

$$
\begin{aligned}
\hat{\boldsymbol{\gamma}}^{OLS} &= \left((\mathbf{X}\otimes\mathbf{W})'(\mathbf{X}\otimes\mathbf{W})\right)^{-1}(\mathbf{X}\otimes\mathbf{W})'\mathbf{y} \\
&\overset{(4.15)-(4.17)}{=} \left((\mathbf{X}'\mathbf{X})^{-1}\otimes(\underbrace{\mathbf{W}'\mathbf{W}}_{\mathbf{I}_n})^{-1}\right)(\mathbf{X}'\otimes\mathbf{W}')\mathbf{y} \\
&\overset{(4.17)}{=} \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{W}'\right)\mathbf{y} \\
&\overset{(4.17)}{=} \left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\otimes\mathbf{I}_n\right)(\mathbf{I}_r\otimes\mathbf{W}')\mathbf{y} \\
&= \left(\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{p\times r}\otimes\mathbf{I}_n\right)\begin{pmatrix}\mathbf{W}'\mathbf{y}_1\\ \vdots \\ \mathbf{W}'\mathbf{y}_r\end{pmatrix}_{nr\times 1} \qquad (4.19) \\
&\overset{(4.18)}{=} \operatorname{vec}\left(\underbrace{(\mathbf{W}'\mathbf{y}_1|\cdots|\mathbf{W}'\mathbf{y}_r)}_{n\times r}\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right) \\
&= \begin{pmatrix}(\mathbf{W}'\mathbf{y}_1|\cdots|\mathbf{W}'\mathbf{y}_r)\ (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})(\cdot,1)\\ \vdots \\ (\mathbf{W}'\mathbf{y}_1|\cdots|\mathbf{W}'\mathbf{y}_r)\ (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})(\cdot,p)\end{pmatrix} = \begin{pmatrix}\hat{\boldsymbol{\gamma}}_1^{OLS}\\ \vdots \\ \hat{\boldsymbol{\gamma}}_p^{OLS}\end{pmatrix}
\end{aligned}
$$

Note that the vector in Eq. (4.19) contains the 3d wavelet transforms, $\mathbf{W}'\mathbf{y}_i$, $i=1,\ldots,r$, of the raw images, thereby "eliminating" $\mathbf{W}$. If thresholding was directly applied to these wavelet coefficients, the proposal would work in analogy to the approach e. g. considered in Hahn et al. (2001), where smoothing comes prior to tensor estimation. In Eq. (4.19), the tensor model is however fitted in the wavelet domain. Hence, model (4.14) could alternatively be stated:

$$\mathbf{y}_{\mathrm{W}} = (\mathbf{X}\otimes\mathbf{I}_n)\boldsymbol{\gamma}+\boldsymbol{\varepsilon}_{\mathrm{W}},$$

with $\mathbf{y}_{\mathrm{W}} = (\mathbf{I}_r\otimes\mathbf{W}')\mathbf{y}$ and $\mathrm{Var}(\boldsymbol{\varepsilon}_{\mathrm{W}}) = \mathrm{Var}\left((\mathbf{I}_r\otimes\mathbf{W}')\boldsymbol{\varepsilon}\right) = \sigma^2\mathbf{I}_{rn}$.
The first approximations $\hat{\boldsymbol{\gamma}}_j^{OLS}$ to the amplitudes of the space-varying coefficient surfaces are independently thresholded using rule δ and cut-off value λ_j, $j=1,\ldots,p$:

$$\hat{\boldsymbol{\gamma}}_{\lambda_j,j} = \delta_{\lambda_j}\left(\hat{\boldsymbol{\gamma}}_j^{OLS}\right).$$

Finally, smooth estimates

$$\hat{\boldsymbol{\beta}}_j = \mathbf{W}\hat{\boldsymbol{\gamma}}_{\lambda_j,j}$$

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

can be obtained for the varying coefficient surfaces and thus the 3d diffusion tensor elements. The thresholding mechanism certainly represents the crucial part of the proposed wavelet based SVCM since it determines the amount of shrinkage. For example, a surface specific universal threshold could be applied in combination with either hard or soft thresholding:

$$\lambda_j^{UNIV} = \sigma_j \sqrt{2\log(n)}.$$

According to Donoho and Johnstone (1994), the noise level σ_j is estimated on the basis of the contributions from the $n/8$ high frequency components. Assume that these elements of the wavelet transform are stored on positions $1, \ldots, n/8$ in vector $\mathbf{W}'\mathbf{y}_i$. Thus each of the first $n/8$ rows of $\left(\mathbf{W}'\mathbf{y}_1 | \cdots | \mathbf{W}'\mathbf{y}_r\right)$ contains the weights associated with a specific high frequency wavelet in the r raw images. The final noise estimation would then involve the linear combinations of those r weights with the jth column of $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ for the given totality of high frequency wavelets. See Section 4.1.4 for details on alternative methods.

In conclusion, the outlined SVCM permits to expect reasonable if not even better results than generated by the B-spline based SVCM. Yet, first endeavors point at concealed problems. A thorough implementation and investigation of performance is deferred to future research.

Acknowledgements

I am grateful to Thomas Kneib (Ludwig-Maximilians-University, Munich, Germany) for helpful comments, to Peter Müller (University of Texas, Houston, USA) for his initial support and to Philipp Sämann (Max Planck Institute of Psychiatry, Munich, Germany) for advice on neuroscientifical applications. The human DTI data stems from my research period at the Max Planck Institute of Psychiatry (Munich, Germany). This study was kindly funded by the collaborative research center SFB386 *Statistical Analysis of Discrete Structures* as part of the German Science Foundation (D. F. G.).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

5 Fiber tracking on the basis of SVCM estimates

As mentioned previously, the term diffusion refers to the random mobility of molecules due to their thermal energy. DTI is specifically sensitized to diffusion of hydrogen, the most frequent and omni-present element in the human body. The mobility of water molecules is mathematically described by a 3d Brownian process with location dependent diffusion matrix $\mathbf{D}(x_t)$ at x_t:

$$dx_t = \mathbf{D}^{\frac{1}{2}}(x_t)dw_t, \tag{5.1}$$

where $t \geq 0$ is "time" after starting from a seed point $x_0 \in \mathbb{R}^3$, and w_t is the realization of a 3d Wiener process (compare also remarks to Eq. (3.2)). Importantly, in biological tissues, various natural barriers such as cell membranes restrict the water molecules in their random displacements. Such restrictions may be spatially homogeneous, leading to overall limited, but still isotropic diffusion, or inhomogeneous, leading to anisotropic diffusion. In the latter case, there is a strong directionality of diffusion as e. g. in the area of white matter fiber tracts in the brain where multiple axons and myelin sheaths are lying parallel to each other. By this alignment densely packed fiber bundles are formed in which any molecular movement perpendicular to the main direction is hampered. In principle, this holds in grey matter, too. However the extent of parallel fiber pieces there is much smaller than the voxel dimension, so that the effect is averaged away by many more or less random fiber orientations. Note that the statistical entity of a fiber is not identical with the anatomical entity of a fiber, i. e. the axon of a neuron: Within a typical voxel from MR imaging ten to hundred thousands of fibers are shaping a statistical average of fiber directionality that itself is then used to detect probable estimates of fiber courses.

The key idea underlying any tracking algorithm is two associate the local fiber direction

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

with the principal eigenvector, ev, i. e. with the largest axis of the local diffusion ellipsoid and hence with locally predominant architecture. The actual fiber propagation is usually based on the Frenet equation (see e. g. Basser et al., 2000), which describes the development of a continuous 3d curve, $x(t)$, parameterized in t:

$$\frac{dx(t)}{dt} = \mathbf{v}(t). \tag{5.2}$$

Following the tracking principle, the eigenvector $ev(t)$ replaces the tangential vector $\mathbf{v}(t)$. The differential equation can be solved numerically, provided a starting point, say $x(0)$ is defined. For example, the well-known method of Euler evolves a Taylor series up to the second term, yielding a neighboring point:

$$x(1) = x(0) + \mathbf{v}(0) = x(0) + ev(0).$$

Iterated application of Euler's method leads to a 3d space curve, $x_0, x_1, \ldots, x_t, \ldots$, interpretable as neuronal fiber. For generalization, the procedure is endowed with a relaxation parameter α, which determines the step size and hence the continuity of the constructed fiber tract:

$$x(t) = x(t-1) + \alpha ev(t-1). \tag{5.3}$$

In practice, however, the eigenvector is observed with noise and has to be derived from the local tensor estimate. Therefore, a fiber that merely bases on the Frenet framework would appear quite wiggly. Several smoothness restictions are imposed, e. g. limited fiber kinking. Additional improvement is achieved by regularization of the tensor field prior to tractography. First successful tractography was reported by turn of the millennium. Current fiber tracking techniques range from simple heuristic techniques to (deterministic) line propagation algorithms (Conturo et al., 1999; Basser et al., 2000; Tench et al., 2002), probabilistic algorithms (Hagmann et al., 2003; Behrens et al., 2003) and sophisticated Bayesian approaches involving spatially complex Markov random field prioris (Poupon et al., 2000; Poupon et al., 2001; Lu et al., 2006). Review articles on tractography principles, strategies and pitfalls are available (Lori et al., 2002; Mori and van Zijl, 2002). A steady update of contributions to the large variety of tractography methods can be found on the web (http://www.cs.brown.edu/research/vis/resources/lists/tractography.html). The next section serves the description of the algorithm used throughout this chapter.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

5.1 Tractography using linear state space models

Instead of iteratively solving the differential equation (5.2), fiber tracts can alternatively be related to a random walk in space. The trajectory errors happen to be spatially dependent and inhomogeneous as described by the estimate of local diffusion covariance. Gössl et al. (2002) embedded a discretized version of the Brownian process (Eq. (5.1)),

$$x_t = x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \mathbf{D}(x_{t-1})) \tag{5.4}$$

as transition equation for the latent curve x_t in a linear state space model, observed with noise:

$$y_t = x_t + \eta_t, \quad \eta_t \sim N(0, \sigma^2 I).$$

In contrast to a conventional linear state space model, observations $\{y_t, t = 1, \ldots, T\}$ have to be generated from the diffusion tensor data. The noisy (pseudo-) observations y_t of x_t can be sequentially obtained from an equation motivated by the Euler method (compare Eq. (5.3)):

$$y_t = \hat{x}_{t-1} + \alpha ev_{t-1}, \quad t = 1, 2, \ldots$$

with $\hat{x}_{t-1}$ the estimated current state of curve, ev_{t-1} the principal eigenvector of the tensor $\mathbf{D}(\hat{x}_{t-1})$, and α a step size parameter. Now the linear state space model can be fitted using Kalman filtering and smoothing. The first order random walk in Eq. (5.4) plays the role of a difference penalty aiming at regularized estimation of x_t. Compared to other tractography algorithms, the information of local diffusivity is taken into account twice: partially by the principal eigenvector for the generation of pseudo-observations and completely by the Kalman gain used in the correction step of the Kalman filter. The recursions of either Kalman filter and smoother can be found e. g. in Fahrmeir and Tutz (1994, Chap. 8.1.2).

Since MR raw data is contaminated with noise, smoothing of the diffusion weighted images or alternatively of the tensor estimates $\hat{\mathbf{D}}$ at original resolution is required for proper track reconstruction. The SVCM approach developed in Chapter 3 allows to spare such a preprocessing step. In addition, a constraint for avoiding too wiggly and unplausible, highly curved fibers is introduced by angle restriction. Specifically, if two subsequent directions enclose an angle more than 70°, the tracking algorithm assumes a fiber crossing

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

and continues in the direction from the previous position. As a consequence, recursive application of the (modified) Kalman filter and smoother was shown to provide fairly smooth estimates $\{\hat{x}_t, t = 0, 1, ...\}$ of unobserved fiber trajectories (Gössl et al., 2002). Note that the tracking accuracy depends on the step size parameter α. By default, quadratic interpolation is employed to assess $\mathbf{D}(\hat{x}_t)$. When tractography is linked to the SVCM approach, the tracking algorithm needs to interact permanently with the estimated B-spline coefficients. For each step, the relevant basis functions are determined in dependence on the knot partition and weighted by their corresponding amplitudes. A fiber trajectory is stopped if either

- a region of low anisotropy is entered where fiber occurence is less likely or even implausible or
- the fiber reaches a still reasonable maximum length or
- the fiber is running out of the imaging volume.

Concerning the technical framework, the tracking algorithm is implemented in C++, while SVCM estimation is performed in R (R Development Core Team, 2006) as detailed in Section 3.9. Visualization of the fiber results can be achieved with the interactive tool TRAX (Pütz et al., 2003), written in the commercial programming language IDL (Interactive Data Language, Research Systems Inc.).

5.2 An example: synthetic data

For a proof of concept, the original tracts as depicted in Fig. 4.13 are constructed from a region of interest (ROI) containing two voxels, one positioned in the inner spiral and the other voxel in the outer spiral in a central axial slice. These are superimposed on the tracking results as red-colored blobs (Fig. 5.1). The algorithmic parameters comprise five seed points per ROI-voxel, 300 tracking steps in each direction along the starting vectors, a step size of $\alpha = 0.5$, and a binary anisotropy cut-off set to FA $= 0.03$. At first sight, the anisotropy threshold may appear exceptionally low compared to the simulation design with FA around 0.4 in spiral voxels and zero in the background. This decision is

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

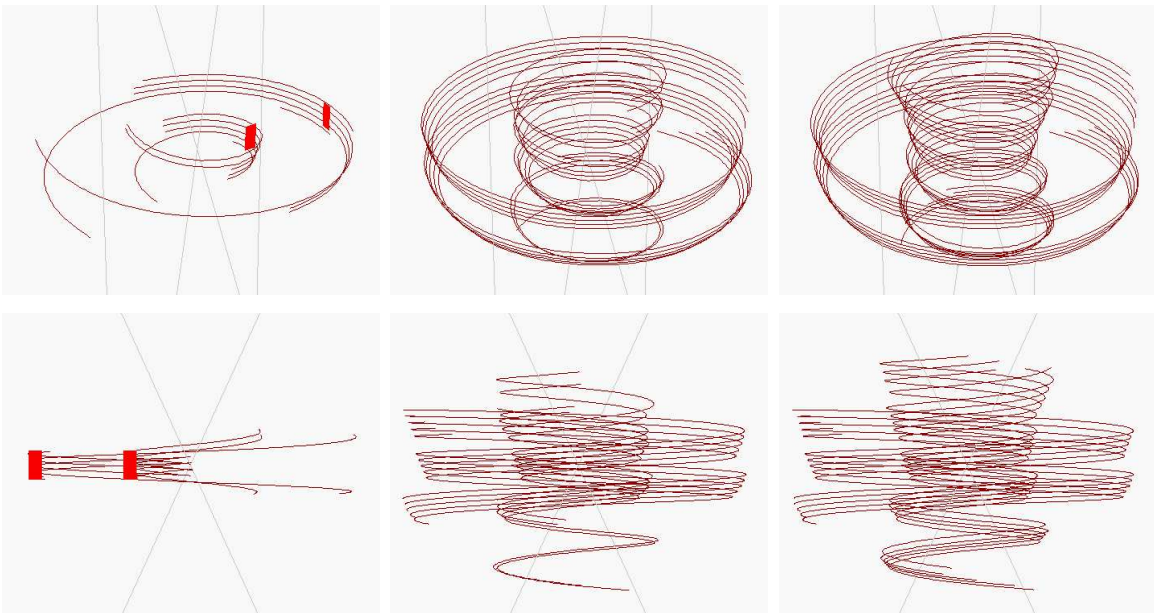


Figure 5.1: Left to right: Tractography results on the basis of the template tensor field without posthoc regularization and with smoothing using a Gaussian kernel of FWHM = 0.75 and FWHM = 1.25 voxels, respectively. The tensors were quadratically interpolated where necessary. ROI voxels are uniquely indicated in red.

subject to some arbitrariness: In choosing 0.03 the fact is accounted for that, in human brain data, the highest physiological FA values range above 0.95. At the same, a typical cut-off in human brain data is 0.2, i. e. about 20% of the maximum value. See below for the explanation of the final threshold.

Importantly, the simulation model in its functionality as template exposes an on/off degree of anisotropy between the two artificial tissue types. During the tracking process the growing fiber pathway drifts to the outer border of the spiral and, at some point, falls off into fully isotropic background (Fig. 5.1, left). This effect can be mitigated by the application of a Gaussian kernel smoother prior to tractography. With increasing kernel size, more and more revolutions of the original bundle are detected (Fig. 5.1, center and right). The ratio of breakups due to a failure of the FA threshold and due to reaching the maximum number of steps reduces from 1 to 0.4 and 0.1 in case of a kernel's FWHM of 0, 0.75 and 1.25 voxels, respectively.

"Drifting" particularly occurs after amplified smoothing: The perspective as well as the front view exhibit a vertically waisted form of both bundles (Fig. 5.1, right). The reasons for this drifting remain unclear. One possibility is that smoothing affects not only the

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

tensor shape of background voxels but also changes the spatial orientation of spiral tensors towards relaxed coherence with the underlying geometry. Also, the value of the step size could be chosen too large. Yet, the most plausible explanation arises from the tracking algorithm itself. Since the entire procedure of recursive Kalman filtering and smoothing can be understood as penalized non-parametric regression (Fahrmeir and Tutz, 1994), underestimation of high curvature could be expected from the (non-adaptive) smoothing scheme. Specifically in the monotonically rotating spiral model bias is steadily accumulated, eventually causing the drift phenomenon.
Furthermore, both artificial fiber bundles expand about two voxels in diameter, not leaving much room for the centrifugally developing pathways. In an overall conclusion from the above findings, the simulation model is suitable to only a limited extent to reliably evaluate the linkage of SVCM methodology to tractography.

The following analyses concentrate on the inital question of whether tractography can be improved by the unified framework for estimation, regularization and interpolation of the 3d tensor field. Fiber construction is examined on the basis of

- standard voxelwise regression without or with subsequent Gaussian smoothing and quadratic interpolation, using a kernel size set to FWHM = 0 (ST1), 0.75 (ST2) and 1.25 voxels (ST2+),
- sequential SVCM with first order difference penalties, a global tuning parameter and either linear (DXglob-lin) or quadratic tensor product B-splines (DXglob-quad), and
- standard voxelwise regression with appended denoising using D2 wavelets and level dependent SURE shrinkage as well as quadratic interpolation (sSURE-D2).

All other algorithmic settings are left untouched. The small anisotropy threshold of 0.03 accounts for the fact that some methods already operate as smoothing tool. As a consequence, the FA values e. g. derived from SVCM tensor estimates are on average 30% lower than those from ST1 tensors.

As evident from Fig. 5.2, the SVCM approach leads to tracking results of intermediate quality compared to the results of the standard pipeline: Neither severely erroneous tracts occur nor is the maximum possible fiber length achieved. While ST1 based tractography causes 18 out of 20 fiber endings due to the anisotropy threshold, this stopping criterium

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

applies only in 13 cases of ST2 based tractography but in all cases of both SVCM settings. The output of DXglob-lin shows slightly more mistaken fibers of overall less trajectories than the output of DXglob-quad.

With tracking on the template tensor field in mind (Fig. 5.1), the SVCM performance does not seem immediately disillusioning anymore. The suspicion of unsufficient smoothing arises for both SVCM approaches. Figure 5.3 serves to clarify the discrepancy in rating the SVCM superior to ST1 if tensor elements are considered (see MSE boxplots in Section 3.4) and inferior to ST1 if tracking results are compared. For this purpose, tractography was conducted on the basis of amplified smoothing of the standard tensor field (ST2+). The corresponding graphics (Fig. 5.3, top) reveal almost the maximum detectable fiber bundles as only two of 20 half-fibers do not pass the FA threshold. In addition, the use of the DXglob-lin and DXglob-quad method for pure tensor estimation is investigated appending Gaussian regularization (FWHM = 0.75 voxels) and quadratic interpolation in analogy to the standard serial procedure. Since the SVCM variants consider spatial correlation already to a certain amount, the respective tracking results in Fig. 5.3 are smoother than those from ST2 in Fig. 5.2 but not as smooth as those from ST2+. In particular, the fiber density in the top third of the front views allows to rank ST2+ superior to serial DXglob-quad and serial DXglob-lin with respect to the most completely fiber findings. Among the approaches involving linear and quadratic B-splines, seven and nine of 20 half-fibers, respectively, interrupt because of limited anisotropy.

The Gaussian smoothing step can be replaced by wavelet shrinkage. Compared to ST2 based tractography, results from wavelet based tractography appear quite poor (Fig. 5.4, central rows). They do however convince in terms of reliability if opposed to ST1 based tractography (top row). Moreover, there is a strong similarity to the tracts gained from the original tensor field (Fig. 5.1, left). This means, that the application of wavelets on top of ST1 works rather in a denoising than in a smoothing fashion. Interestingly, additional Gaussian smoothing (Fig. 5.4, bottom row) does not bring up as long fiber pathways as present after Gaussian smoothing of the original (Fig. 5.1, center) or of the voxel-by-voxel estimated tensor field (Fig. 5.4, ST2). In this latter case of modified wavelet based tractography, there are still 16 breakups owing to the anisotropy threshold.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

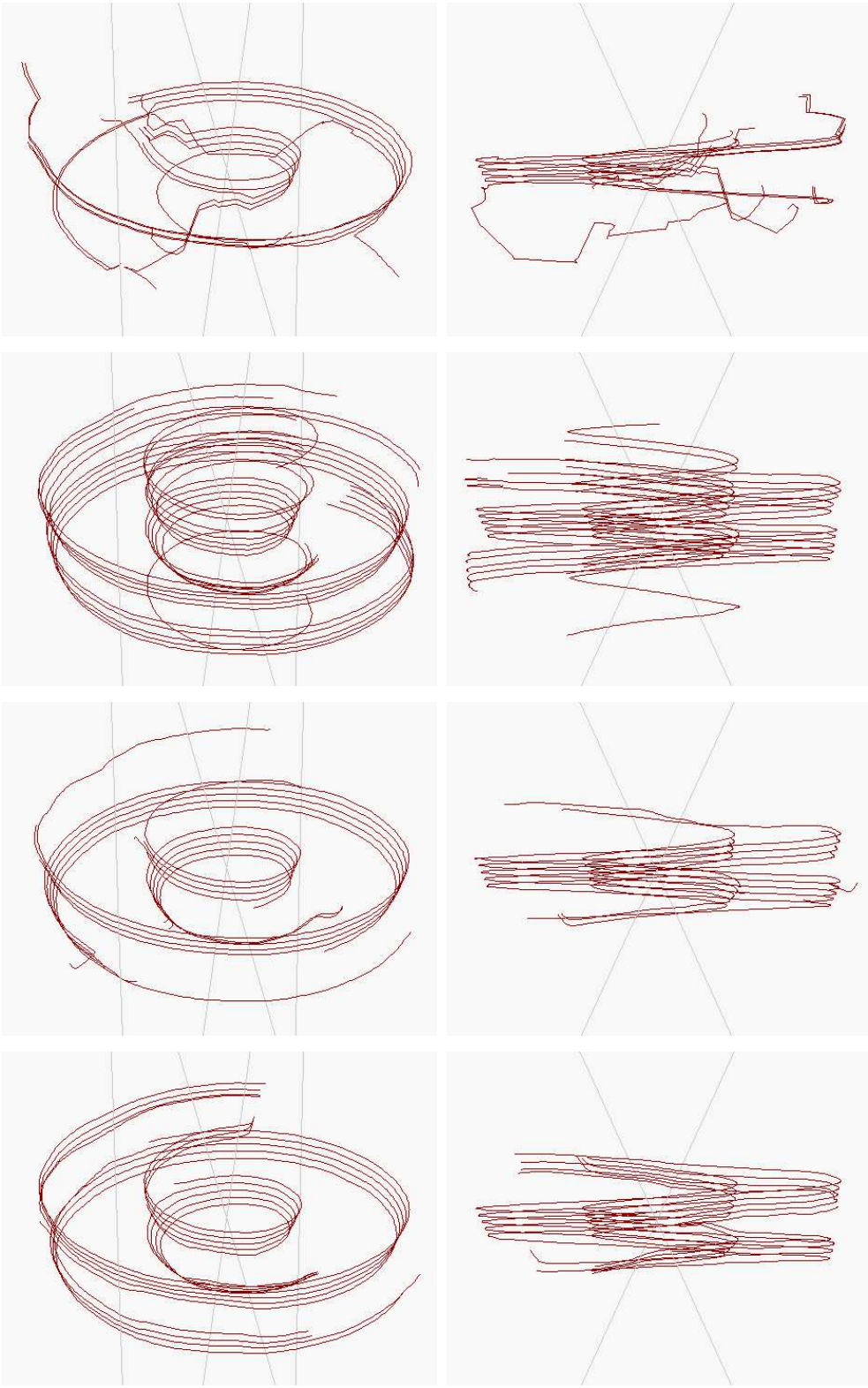


Figure 5.2: Top to bottom: Constructed fibers using standard data processing without and with Gaussian kernel smoothing (FWHM = 0.75 voxels) prior to tractography, and SVCM approaches with linear and quadratic B-splines.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

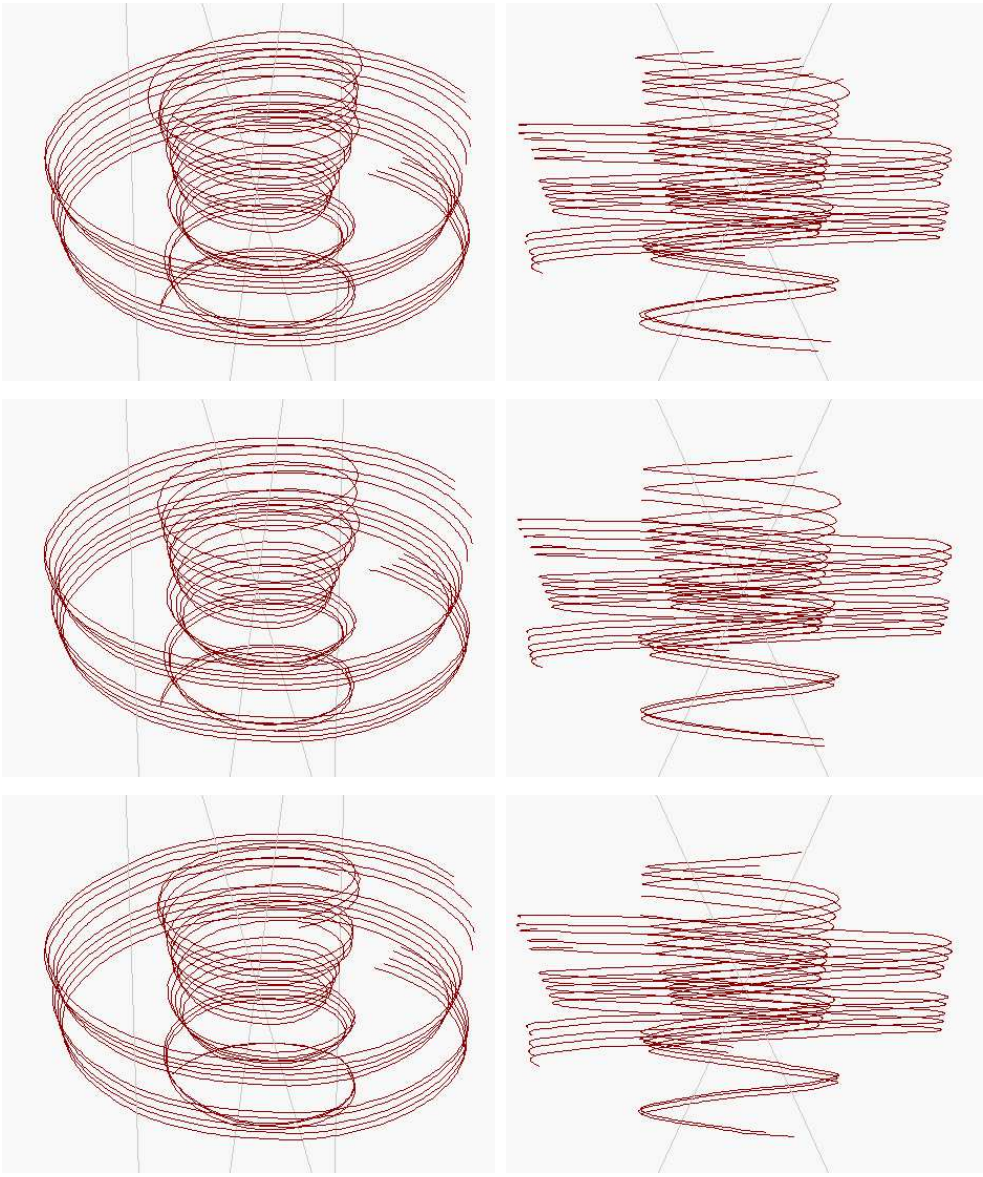


Figure 5.3: Tractography results gained on the basis of ST2+ (top) and alteratively, using linear (center) and quadratic (bottom) B-splines purely for tensor estimation, followed by additional Gaussian smoothing (FWHM = 0.75 voxels). Quadratic interpolation is employed in all approaches.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

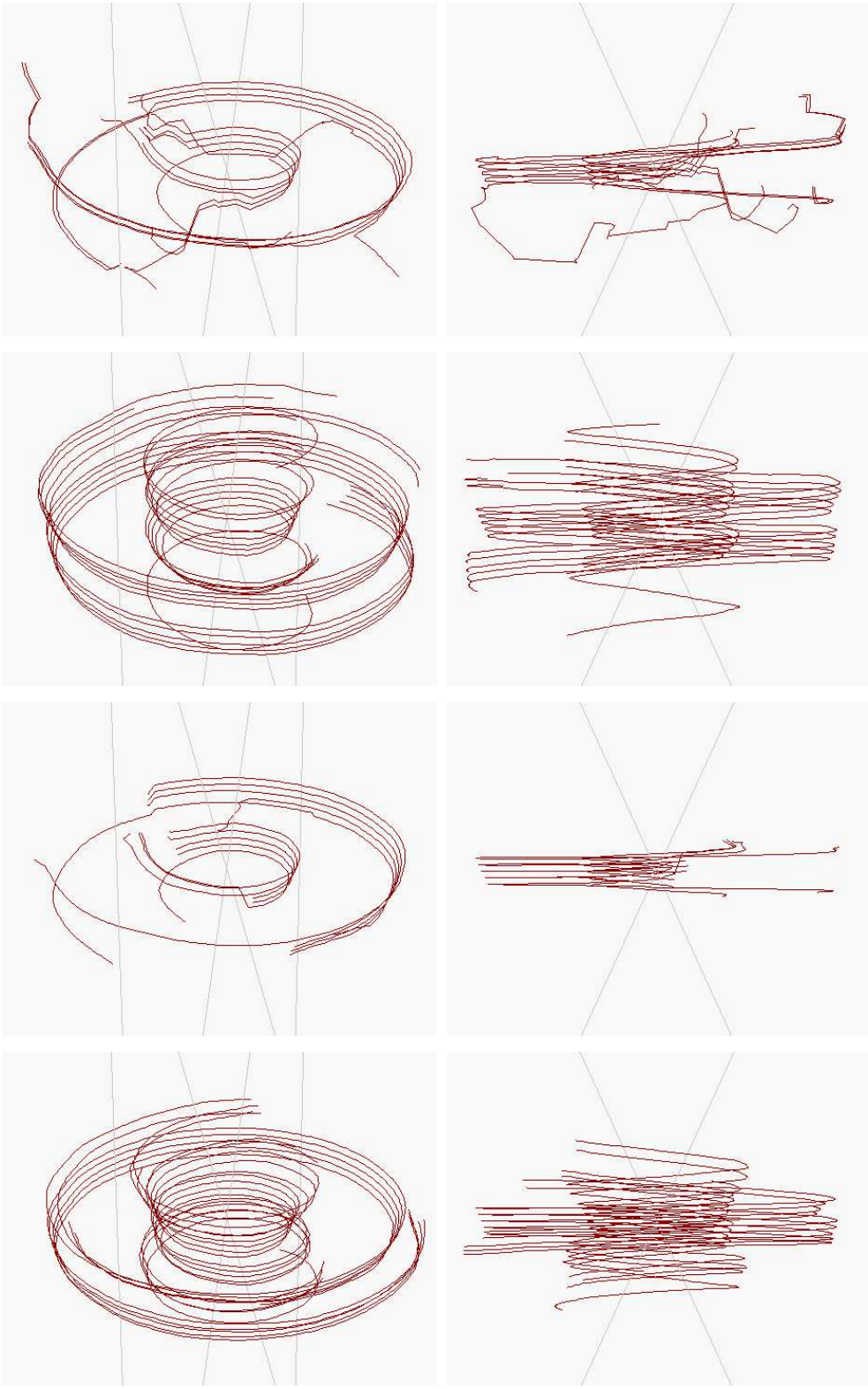


Figure 5.4: Top to bottom: Tractography results generated on the basis of ST1, ST2, sSURE-D2 and sSURE-D2 with additional Gaussian smoothing (FWHM = 0.75 voxels). Quadratic interpolation is employed in all approaches.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

5.3 An example: human data

To check the plausibility of the algorithm for human brain data, reconstruction is exemplified by means of some fiber bundles of known clear anatomy and of low interindividual variability. A popular object is the pyramidal tract, a major pathway running bilaterally from the motor cortex to the spinal cord. In this study, the pyramidal tract is to be recovered from two starting regions which were drawn on two adjacent axial slices of the FA-weighted direction map. The final ROI comprises 45 voxels in total and will be depicted as blue-colored blobs. Methodological investigation focusses on the same approaches that have already been examined with synthetic data. Tracking parameters are adjusted to the conditions of human data by changing to 500 steps and a common anisotropy threshold of 0.15. As before, five starting points are uniformily distributed in each voxel of the region of interest.

Figure 5.5 displays the tracking outcome achieved by ST1, ST2 and the two SVCM settings, DXglob-lin and DXglob-quad. As anticipated from the spiral, skipping the regularization step leads to generally wiggly pathways, a considerable amount of which appears incorrect (Fig. 5.5, top). The results are found to be partly divergent from known anatomy as particularly obvious from the front and left views. For example, there should be no horizontal fibers for the given ROI. The connections between the two pyramidal bundles presumably belong to another formation known as corpus callosum. If Gaussian kernel smoothing is applied, the partial structure of the corpus callosum is even clarified (Fig. 5.5, second row). This means that smoothing does not prevent the algorithm from hopping on to neighbouring fiber bundles. Yet, an overall impression of smooth alignment is gained as appropriate from the biological point of view.
Both fiber bundles loose somewhat of this smoothness if recovered with DXglob-lin based tractography (Fig. 5.5, third row). Apart from a single spin-off running towards the left frontal lobe, only a few fiber artifacts emerge. In fact, DXglob-lin is the only method that restrains the tracking algorithm from entirely brigding along the corpus callosum. In contrast, DXglob-quad seems to produce a tensor field of unsufficient edge-preservation, eventually resulting in too liberal tractography (Fig. 5.5, bottom). As it can be judged from the smooth loops, regularization is not the primary concern with this approach.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

With regard to the statistics, around 82–93% of the fiber endings happen to be outside the imaging volume. Regions of low anisotropy are entered in 18%, 7%, and 13% of the half-fibers constructed on the basis of ST1, ST2 and the SVCM settings, respectively.

Tractography with enlarged kernel size, namely ST2+, can slightly improve fiber smoothness (Fig. 5.6, top). A more prominent role plays the reduced number of erroneous pathways. Results from ST2+ are hardly distinguishable from those found by the pipeline of DXglob-lin for mere tensor estimation with appended Gaussian smoothing of default kernel size (FWHM = 0.75 voxels) and quadratic interpolation (center). Tracking lines from the DXglob-quad analogon are also running in great regularity apart from a few outliers (bottom). The FA threshold causes 6% of the fiber endpoints regardless the method used.

At first sight, the denoising effect of wavelet filtering is not as strong as in the synthetic data as evident by comparison to ST1 results (Fig. 5.7, first and third row). After additional smoothing, however, the fibers align in a well-organized fashion and exhibit less artifacts, also than ST2 (second and fourth row). This again corroborates the denoising theory inherent to wavelet application. There are 3% less fibers reaching the borders of the imaging volume when tractography involves both wavelet and Gaussian smoothing.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

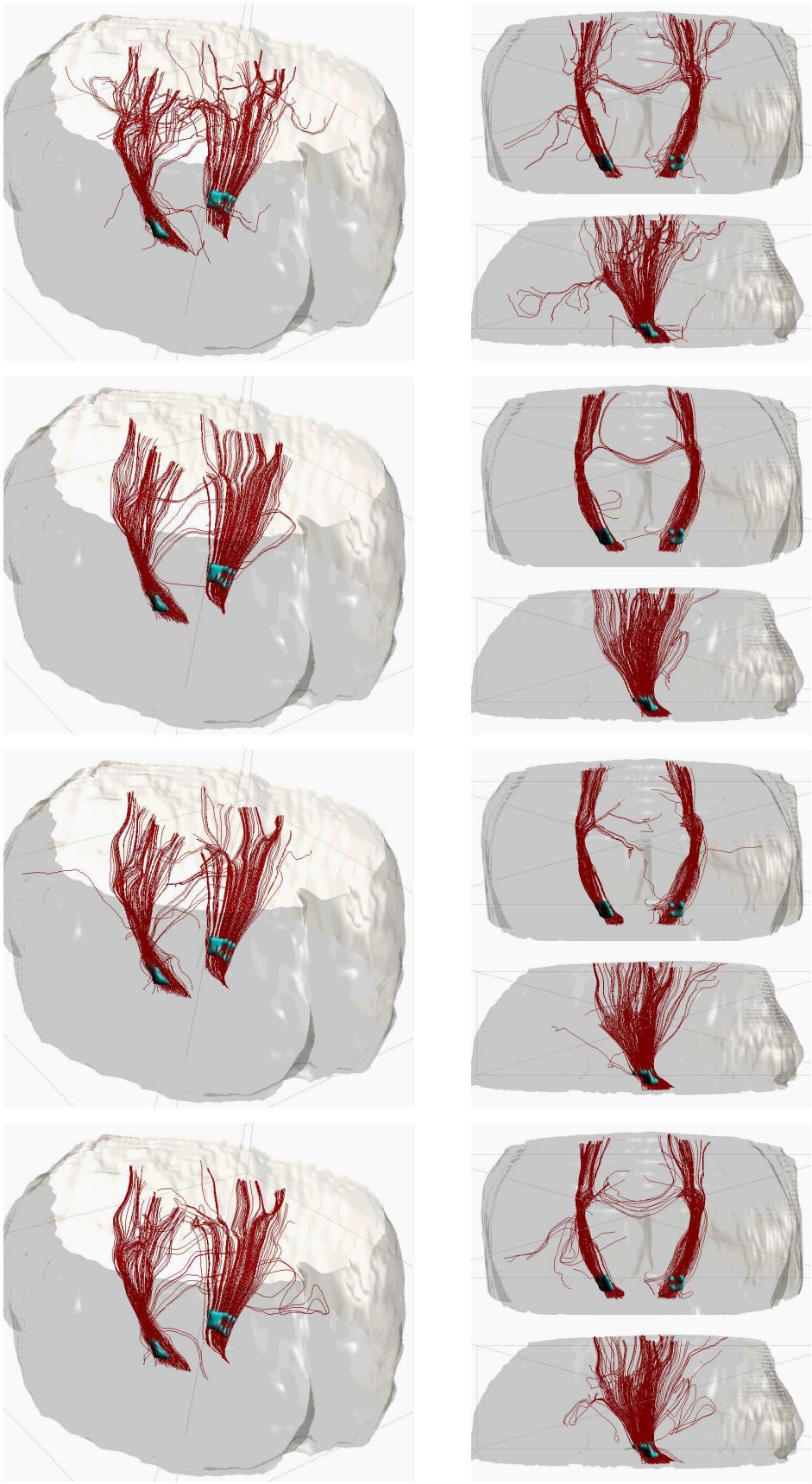


Figure 5.5: Top to bottom: Perspective, front and left view of the pyramidal tract as reconstructed using ST1, ST2, DXglob-lin and DXglob-quad.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

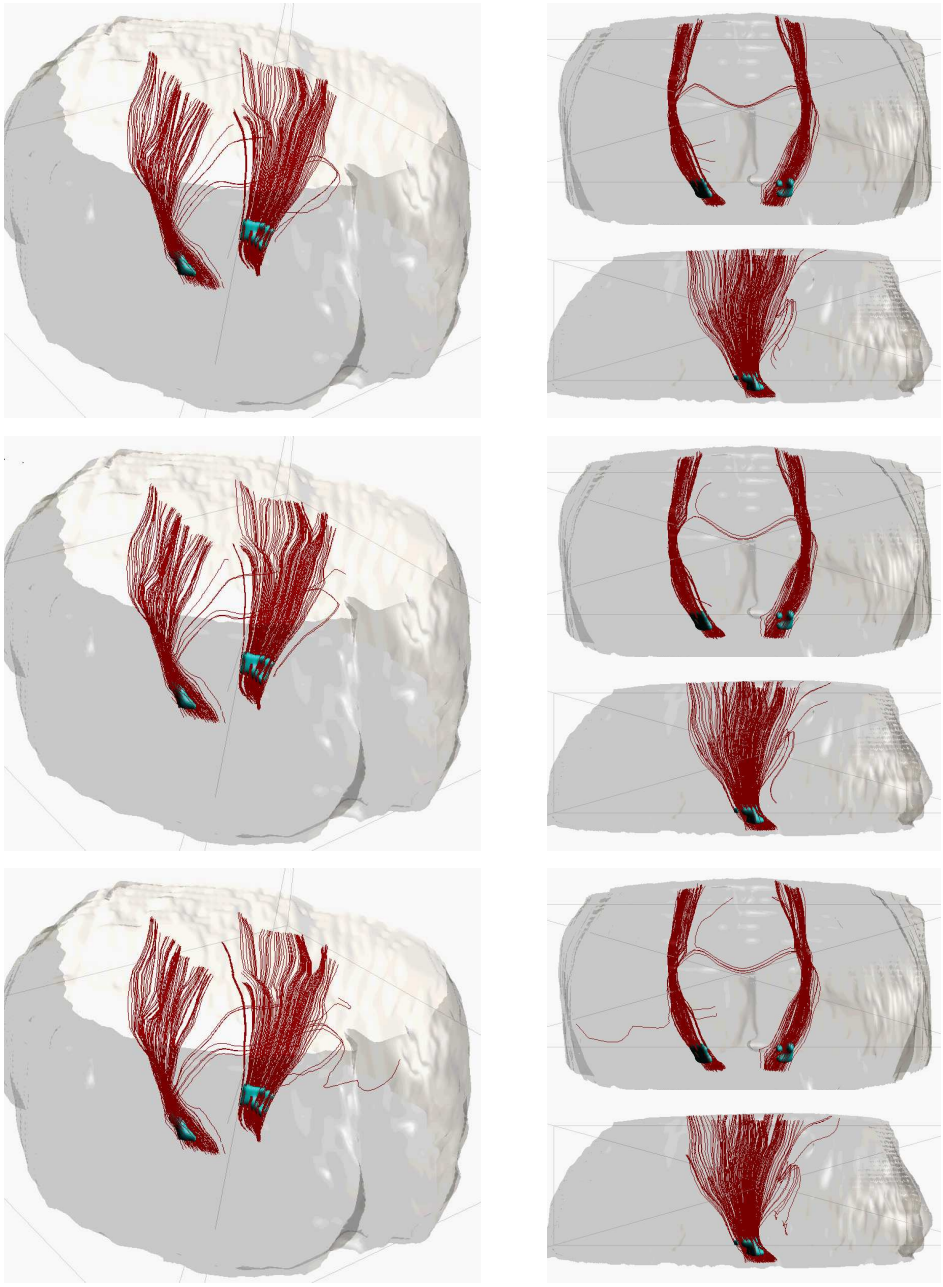


Figure 5.6: Top to bottom: Perspective, front and left view of the reconstructed pyramidal tracts as constructed using ST2+, SVCM with linear and quadratic 3d B-splines, respectively, followed by Gaussian kernel smoothing (FWHM = 0.75 voxels) and quadratic interpolation.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

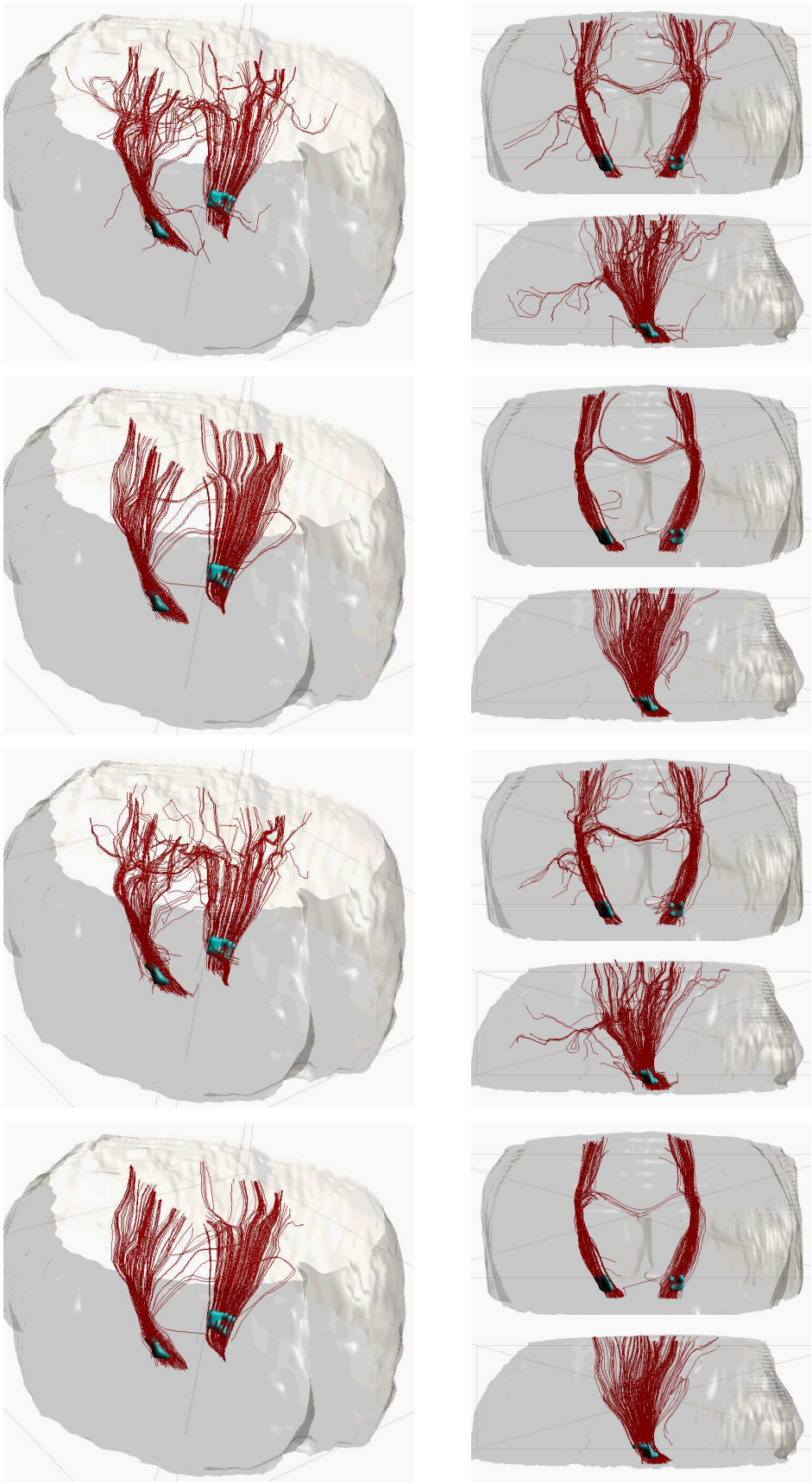


Figure 5.7: Top to bottom: Perspective, front and left view of the pyramidal tract as reconstructed using ST1, ST2, sSURE-D2 without and with Gaussian smoothing (FWHM = 0.75 voxels).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

5.4 Discussion

Tractography on the basis of linear state space models (Gössl et al., 2002) was successfully conducted for different approaches of DTI data processing, recovering simulated fiber geometry and plausible real fiber tracts in the human brain. When assessed with regard to their validity, tracking results gained from the novel SVCM framework without posthoc Gaussian smoothing were about equivalent to the reference processing pipeline comprising voxelwise tensor estimation, Gaussian kernel smoothing and quadratic interpolation. Wavelet filtering after voxelwise tensor estimation could not improve the fiber tracking results unless supplemented by Gaussian smoothing. In total, sufficient regularization seems to be a key factor for satisfactory fiber construction as strongly indicated by comparative tractography carried out at different degrees of smoothing, either as part of the SVCM or as subsequent step. However, the upper limit of regularization that does not impair reliable fiber reconstruction has to be defined from further investigations. The study also revealed that in the context of tractography, the application of wavelets requires careful selection of the thresholding technique and the choice of an appropriate wavelet class.

The discrepancy between the conclusions based on the MSE boxplots on the one hand (Sections 3.4 and 4.3.4) and the conclusions based on the visualization of reconstructed fibers on the other hand is notable and can be elucidated as follows: While the MSE of the tensor estimates represents a measure of local performance, the smoothness of the whole tensor field as the basis of fiber reconstruction is only accessible on the spatial level. According to the variance-bias decomposition, a high MSE value may primarily be caused by high variability or high bias of the estimator, i. e. by over- or underfitting, respectively. An MSE averaged over all voxels (AMSE) does therefore not allow to differentiate between a generally undersmoothed and a generally oversmoothed tensor field. If, however, the MSE is built on a voxel-by-voxel basis over several simulation runs (VMSE), the resulting error map provides visual information about the edge-preserving and smoothing properties of the estimator depicted on the VMSE ratio maps of Fig. 3.8 and Fig. 4.17. The developed B-spline based SVCM led to an estimator that can deal with sharp edges but at the expense of regularity, that is, the behaviour is contrary to the standard approach which involves Gaussian kernel smoothing. Local adaptivity, as an

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

expression of simultaneously minimized variance and bias, was sought to be improved by the use of wavelet filters. In confirmation of the multiscale property, wavelet filtering rendered the tensor elements of synthetic data into surfaces of high visual quality (Fig. 4.18). Importantly, this impression was in good consistency with the results from the AMSE and VMSE boxplots (Fig. 4.16). However, when the methods were applied to real data and diffusion ellipsoids were visualized, wavelet filtering could not corroborate its superiority (Fig. 4.20). Intuitively, with respect to fiber tractography, more smoothing is desired not only for realistic, complex fiber courses, but also for the reconstruction of simple c-shaped fiber bundles. Also, a more appropriate measure of local adaptivity is urgently required, since the L_2 error is not always adequate for visual interpretation (Härdle et al., 1998, pp. 139–146).

Final conclusions on the overall usefulness of the SVCM approach for fiber tracking cannot be drawn at this stage. First, other simulation models should be considered to infer on preprocessing quality and eventually tracking performance due to the manifold reasons already discussed in Section 3.6. One drawback worth mentioning was the construction of a relatively thin fiber bundle in combination with a fully isotropic background. Despite smoothing can alleviate this restriction, the final result images suggested that this may have impeded fiber propagation due to the intrinsic stop criteria, leading to the false impression of poor performance. The spiral geometry furthermore causes the tracking algorithm to accumulate bias, eventually leading to centrifugally drifting tracts. Second, the reconstruction of selected fiber tracts of the human brain are claimed to comply with the known neuroanatomy. However, more formal confirmation would be desired. Since no gold standard is available and animal models are questionable from an ethical point of view, validation is limited to quasi realistic phantoms except for the recent work of Bürgel et al. (2006) who employed histological sections of postmortem human brain specimen. The phantom models can be generated by averaging or repeatedly acquiring imaging volumes, thereby increasing the signal-to-noise-ratio (SNR), or by contrasting artificially noised data sets with their originals. Detailed instructions how to reasonably simulate neural fiber bundles are given in Leemans et al. (2005) and also Lori et al. (2002). The issue of a quantitative validation has been addressed by Mishra et al. (2006) who propose a similarity coefficient for fiber estimates. Also, the combined application of a quasi

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

gold standard and validation measures allows to objectively compare available tracking algorithms (Lazar and Alexander, 2003; Lu et al., 2006). Finally, the experimental data was chosen to contain only the minimum of gradient directions needed to estimate the diffusion tensor. SNR is accordingly low and smoothing becomes even more important for denoising purposes. It would be interesting to test the new methodology of SVCM on data from more recent DTI acquisition, especially from scanners with higher magnetic field strength that allow for higher resolution and still high SNR.

Two further challenges arise from the limited resolution of DTI raw data and the numerous sources of noise which may lead to ambiguous and deflected direction of the principle eigenvectors: First, a certain degree of uncertainty is always inherent to the estimated major diffusion and thus fiber directionality. To avoid uncontrollable error propagation into derived parameters and fiber tracts, robustness is a desired features of any tracking algorithm. Although most techniques calculate neuronal pathways via point estimation, more recent approaches provide confidence measures of the reconstructed fibers, for example by using bootstrap procedures (Lazar and Alexander, 2005; Jones and Pierpaoli, 2005) or Bayesian methods (Behrens et al., 2003; Lu et al., 2006). In principle, confidence bands are also accessible within the linear state space model, though not yet implemented. Second, the tensor tends to be disc-shaped in regions of low anisotropy. Such geometrical information is ambiguous and may indicate (i) either a voxel of other tissue than white matter, (ii) a voxel partially contaminated by a second tissue type (also referred to as partial volume effect) or (iii) a real underlying fiber branching or crossing. Partial volume effects cannot entirely be resolved since increasing resolution reduces SNR. The mixture of different tissue types, however, can be accounted for in e. g. a weighted tensor model (Basser and Jones, 2002). To face the problem of fiber tract dissection, several approaches may be conceivable. The most obvious solution seems to be the exploitation of a possibly known destination region of a tracking process. This could be incorporated into the existing tracking algorithm of (Gössl et al., 2002) within the framework of a Brownian bridge. Alternatively, anatomical a priori knowledge could be introduced into the tracking algorithm in a Bayesian type of model (Hosey et al., 2005; Behrens et al., 2007).

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

6 Conclusion and perspective

Diffusion tensor imaging (DTI) is an advanced magnetic resonance (MR) imaging technique that exploits the diffusion properties of biological tissue to provide in-vivo insights into tissue ultrastructure. In particular in brain imaging, the directionality of diffusion in white matter permits the reconstruction of fiber bundles with the aim to eventually build up atlases of neuronal connections. However, the price for the high sensitivity of DTI to neurophysiological alterations is paid for by various intrinsic noise components and limited image resolution. Tackling the question of MR data quality and uncertainty in derived diffusion tensor variables, an objective, automated and global quality measure was built on the basis of bootstrap analysis. It proved sensitive to minor changes in noise levels and to motion induced artifacts and, by this, helpful to correct for potential confounds in group comparisons. To improve data analysis in general and in particular DTI tractography, the uncontrolled error accumulation and propagation by sequential DTI data processing needs to be reduced. For this purpose, a space-varying coefficient model (SVCM) was conceptualized which allows for the unified estimation, smoothing and interpolation of the diffusion tensor field. By using penalized B-spline basis functions, the SVCM naturally accounts for spatial correlation and incorporates a certain amount of regularization. The resulting tensor field is continuous, making additional interpolation steps dispensable. The computational burden of the 3d tensor product approach could be mitigated by sequential approximation, that allows for application to real data sets from brain imaging. Validation was based on different MSE analyses of simulated data sets and on visual inspection of tensor estimates from real data. Although the novel processing of DTI data was shown to be superior to the standard voxelwise regression followed by Gaussian smoothing, local adaptivity remains to be refined with respect to the neuroanatomical complexity of the human brain. Wavelet basis functions, which were

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

selected because of their multiscale property, could improve the simulation results when employed only as filtering tool. However, when the full tensor information of real brain data was assessed with special software, on a qualitative level, it turned out that the regularization effect of wavelet filtering seemed unsatisfactory compared with the alternative processing pipelines. The impact of the B-spline based SVCM and wavelet filtering on DTI fiber construction was then directly investigated after implementing both approaches into an existing tracking algorithm. In the case of real data, the constructed fibers were found to be consistent with normal anatomy and were of equivalent or slightly better quality compared to those originating from voxelwise tensor estimation and subsequent Gaussian smoothing. After replacement of the Gaussian smoother by a wavelet filter, evidently less and more wiggly fibers were yielded, somewhat confirming the suspicion of insufficient regularization. Admittingly, the simulation model turned out to deserve modifications to allow for more meaningful conclusions. In summary, a sufficient degree of regularization seems to be of higher importance than edge-preservation for reasonable tract reconstruction.

Apart from the application of MSE based measures prior to tractography, the probability can be derived that a certain estimator is within an acceptable range. On a smooth tensor field, confidence intervals are conceivable at each position, again providing a local measure of uncertainty (Section 3.9.6). Future efforts should therefore be invested in the derivation of simultaneous confidence surfaces and, with respect to tractography, to the derivation of confidence bands for the reconstructed fibers. In principle, it is possible to comply with the latter request within the previously described tracking algorithm that is based on the linear state space model (Section 5.1). Moreover, a quantitative assessment of the reliability of the reconstructed fibers is necessary to compare different processing pipelines and tracking algorithms among each other (Lazar and Alexander, 2003; Lu et al., 2006; Mishra et al., 2006). Here, validated synthetic datasets or histologically validated DTI brain datasets have proven useful (Bürgel et al., 2006). With regard to the reported discrepancy between the regularized estimation of tensor elements and mainly eigenvector based tractography performance, the tensor diagonalization step needs to be investigated in more detail. Apart from an error analysis of the tensor elements and tensor derived scalar measures deviation, also angles between the estimated principle eigenvector

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

and the noise-free, template based eigenvector should be analyzed. To explore the impact of the new DTI methodology on neuroscientifical applications, the proposed methods should also be conducted on the latest generation DTI data from higher magnetic fields.

Methodological extensions have already been anticipated both for the SVCM and the tracking algorithm. The introduction of penalty weights in the B-spline based SVCM as suggested in Section 3.7 seems promising for the improvement of local adaptivity. While being motivated by the same reason, the exchange of the B-spline basis functions with wavelets represents a more substantial challenge. A theoretical model has been proposed (Section 4.6), and from first attempts to implement this model, it may be speculated that by switching the wavelet family or by modifying the thresholding strategy an optimal trade-off between regularization and edge preservation can be found. Such fine tuning of the wavelet based SVCM to achieve smoothness and local adaptivity may also be useful for other types of high-dimensional data, e. g. from functional magnetic resonance imaging, epidemiology or climatology. Also the effective refinement of the resolution of brain images constitutes a valuable functionality, as e. g. benefitial for brain segmentation. Moreover, a continuous tensor field generally opens the opportunity to apply methods of differential geometry. This, for example, could be useful for non-linear spatial transformations as needed for the alignment of an individual DTI data set to a normative atlas space. As pointed out above, the tracking algorithm of Gössl et al. (2002) could be endowed with confidence bands for each fiber of the recovered bundle by exploiting the Kalman filter step. In addition, the reliability of a reconstructed tract could be improved by the inclusion of external knowledge, e. g. known anatomical target regions. This can be achieved by incorporating the end point information into a time-continuous extension of the existing tracking algorithm which is also known as Brownian bridge.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Danksagung

Zu allererst danke ich meinem Doktorvater Herrn Prof. Dr. Ludwig Fahrmeir für sein bedingungsloses Vertrauen und seine ununterbrochene Zuversicht, die mich die Jahre hindurch ermutigten, sowie für seine mir vorbildhafte und unerreichbare Ruhe. Unter die hervorragende Betreuung fällt auch der internationale wissenschaftliche Austausch, den er nachhaltig zu fördern wusste und den er durch sein intensives Engagement im Sonderforschungsbereich 386 grundlegend ermöglichte.

Mein Dank gebührt auch Frau Prof. Dr. Dorothee Auer, die mich nicht nur in die Kunst des präzisen und prägnanten Verfassens wissenschaftlicher Artikel einführte, sondern auch durch aktive Nutzung bereits vorhandener Analyseverfahren und -module sowie Vermittlung der medizinischen Relevanz zum eigenen Beitrag motivierte. Darüberhinaus gewährten mir die ersten 16 Promotionsmonate in ihrer Arbeitsgruppe einen ebenso umfassenden wie wertvollen Einblick in die interdisziplinäre Forschung an der Schnittstelle von Biologie, Medizin, Physik, Psychologie und Statistik.

Philipp Sämann zeigte stets ehrliches Interesse und Verständnis an statistischer Methodenentwicklung und der damit verbundenen Programmierarbeit. In zahlreiche fruchtbare Diskussionen brachte er zudem sein medizinisches Fachwissen ein und sicherte somit die Verbindung zur anwendungsorientierten Forschung. Für seine kontinuierliche, langmütige Begleitung danke ich ihm sehr.

Ganz besonderer Dank gilt auch Thomas Kneib, der diese Arbeit weitgehendst las und mit seinen Kommentaren vor allem zur besseren Verständlichkeit der theoretischen Teile beitrug. Ich danke ihm für die gemeinsamen Jahre im Büro, seine Hilfsbereitschaft und sein Wohlwollen mir und dem Leben gegenüber.

Ferner danke ich Benno Pütz und meinem Bruder Stefan Heim für ihre essentielle Unter-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

stützung bei allen rechnertechnischen Fragen und Problemen. Benno Pütz korrigierte zudem fachkompetent das einführende Kapitel über Magnetresonanzbildgebung, wofür ich ihm herzlich danke.

Abschließend möchte ich allen Kollegen, meiner Familie und vor allem meinen Freunden Thomas Brückner, Philipp Godron, Thomas Kneib, Katja Menzel, Philipp Sämann, Juliane Schäfer, Daniela Stöcker und Tina Wild für Verständnis, Geduld, Zuspruch und Trost danken.

Die vorliegende Arbeit entstand innerhalb eines Kooperationsprojektes zwischen der Arbeitsgruppe Kernspintomographie (Max Planck Institut für Psychatrie, München) und dem Institut für Statistik (Ludwig-Maximilians-Universität München), finanziert durch die Deutsche Forschungsgemeinschaft im Rahmen des Sonderforschungsbereich 386 *Statistische Analyse diskreter Strukturen.*

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Bibliography

Abe, O., Aoki, S., Hayashi, N., Yamada, H., Kunimatsu, A., Mori, H., Yoshikawa, T., Okubo, T. and Ohtomo, K. (2002). Normal aging in the central nervous system: Quantitative MR diffusion-tensor analysis, *Neurobiology of Aging* **23**: 433–441.

Alexander, D. (2004). A comparison of q-ball and PASMRI on sparse diffusion MRI data, *Proceedings of the 12th Scientific Meeting and Exhibition of the International Society of Magnetic Resonance in Medicine (ISMRM'04)*, Vol. 13, Kyoto, Japan, p. 90. **URL:** *http://cds.ismrm.org/ismrm-2004/Files/000090.pdf*

Anderson, A. W. (2001). Theoretical analysis of the effects of noise on diffusion tensor imaging, *Magnetic Resonance in Medicine* **46**: 1174–1188.

Andersson, J. L. R. and Skare, S. (2002). A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI, *NeuroImage* **16**: 177–199.

Arfanakis, K., Gui, M. and Lazar, M. (2006). Optimization of white matter tractography for pre-surgical planning and image-guided surgery, *Oncology Reports* **15**: 1061–1064.

Assaf, Y., Freidlin, R. Z., Rohde, G. K. and Basser, P. J. (2004). New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter, *Magnetic Resonance in Medicine* **52**: 965–978.

Assunção, R. (2003). Space varying coefficient models for small area data, *Environmetrics* **14**: 453–473.

Basser, P. J. and Jones, D. K. (2002). Diffusion-tensor MRI: Theory, experimental design and data analysis – a technical review, *NMR in Biomedicine* **15**: 456–467.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Basser, P. J., Matthiello, J. and Le Bihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo, *Journal of Magnetic Resonance. Series B* **103**: 247–254.

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J. and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data, *Magnetic Resonance in Medicine* **44**: 625–632.

Basser, P. J. and Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI, *Journal of Magnetic Resonance. Series B* **111**: 209–219.

Bastin, M. E., Armitage, P. A. and Marshall, I. (1998). A theoretical study of the effect of experimental noise on the measurement of anisotropy in diffusion imaging, *Magnetic Resonance Imaging* **16**: 773–785.

Bates, D. and Maechler, M. (2006). *Matrix: A Matrix package for R.* R package version 0.9975-6.

Behrens, T. E. J., Johansen-Berg, H., Jbabdi, S., Rushworth, M. F. S. and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?, *NeuroImage* **34**: 144–155.

Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M. and Smith, S. M. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging, *Magnetic Resonance in Medicine* **50**: 1077–1088.

Blu, T. and Unser, M. (2003). A complete family of scaling functions: The (α, τ)-fractional splines, *Proceedings of the 28th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, Vol. VI, Hong Kong SAR, People's Republic of China, pp. 421–424.

Brezger, A., Fahrmeir, L. and Hennerfeind, A. (2005). Locally adaptive gaussian markov random fields in human brain mapping, *SBF386 discussion paper 456*, University of Munich.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Brezger, A. and Lang, S. (2006). Generalized structured additive regression based on bayesian p-splines, *Computational Statistics and Data Analysis* **50**: 967–991.

Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J. M. and Zilles, K. (2006). White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability, *NeuroImage* **29**: 1092–1105.

Cercignani, M., Bozzali, M., Iannucci, G., Comi, G. and Filippi, M. (2001). Magnetization transfer and mean diffusivity of normal appearing white and grey matter from patients with multiple sclerosis, *Journal of Neurology, Neurosurgery and Psychiatry* **70**: 311–317.

Chabriat, H., Pappata, S., Poupon, C., Clark, C. A., Vahedi, K., Poupon, F., Mangin, J. F., Pachot-Clouard, M., Jobert, A., Le Bihan, D. and Bousser, M. G. (1999). Clinical severity in cadasil related to ultrastructural damage in white matter: In vivo study with diffusion tensor MRI, *Stroke* **30**: 2637–2643.

Chang, L. C., Jones, D. K. and Pierpaoli, C. (2005). RESTORE: Robust estimation of tensors by outlier rejection, *Magnetic Resonance in Medicine* **53**: 1088–1095.

Chung, S. W., Lu, Y. and Henry, R. G. (2006). Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters, *NeuroImage* **33**: 531–541.

Clark, C. A. and Le Bihan, D. (2000). Water diffusion compartmentation and anisotropy at high b values in the human brain, *Magnetic Resonance in Medicine* **44**: 852–859.

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., McKinstry, R. C., Burton, H. and Raichle, M. E. (1999). Tracking neuronal fiber pathways in the living human brain, *Proceedings of the National Academy of Sciences of the United States of America* **96**: 10422–10427.

Currie, I., Durbán, M. and Eilers, P. (2006). Generalized linear array models with applications to multidimendional smoothing, *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **68**: 259–280.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dierckx, P. (1982). A fast algorithm for smoothing data on a rectangular grid while using spline functions, *SIAM Journal on Numerical Analysis* **19**: 1286–1304.

Dierckx, P. (1993). *Curve and surface fitting with splines*, Oxford: Monographs on numerical analysis, Oxford University Press.

Dietrich, O., Heiland, S. and Sartor, K. (2001). Noise correction for the exact determination of apparent diffusion coefficients at low SNR, *Magnetic Resonance in Medicine* **45**: 448–453.

Ding, Z., Gore, J. C. and Anderson, A. W. (2005). Reduction of noise in diffusion tensor images using anisotropic smoothing, *Magnetic Resonance in Medicine* **53**: 485–490.

Donoho, D. L. (1992). Interpolating wavelet transforms, *Technical report*, Department of Statistics, Stanford University, CA/USA.
URL: *http://stat.stanford.edu/∼donoho/Reports/1992/interpol.pdf*

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**: 425–455.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Association* **90**: 1200–1224.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Annals of Statistics* **7**: 1–16.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability, Chapman and Hall, London.

Eilers, P. H. C., Currie, I. and Durbán, M. (2006). Fast and compact smoothing on large multidimensional grids, *Computational Statistics and Data Analysis* **50**: 61–76.

Eilers, P. H. C., Heim, S. and Marx, B. D. (2005). Varying coefficient tensor models for brain imaging, *in* A. R. Francis, K. M. Matawie, A. Oshlack and G. K. Smyth (eds), *Statistical Solutions to Modern Problems. Proceedings of the 20th International Workshop on Statistical Modelling (IWSM'05)*, Sydney, Australia, pp. 157–164.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties (with comments and rejoinder), *Statistical Science* **11**: 89–121.

Eilers, P. H. C. and Marx, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression, *Chemometrics and Intelligent Laboratory Systems* **66**: 159–174.

Fahrmeir, L., Kneib, T. and Lang, S. (2004). Penalized structured additive regression for space-time data: A bayesian perspective, *Statistica Sinica* **14**: 731–761.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*, Mathematics and Visualization, Springer-Verlag, New York.

Fillard, P. and Toussaint, N. (2006). *Medical Image Navigation and Research Tool by INRIA (MedINRIA) Tutorial v1.0*, Institut National de Recherche en Informatique at en Automatique (INRIA) - Research Project ASCLEPIOS, Sophia Antipolis, France. **URL:** *http://www-sop.inria.fr/asclepios/software/MedINRIA/*

Frank, L. R. (2002). Characterization of anisotropy in high angular resolution diffusion-weighted MRI, *Magnetic Resonance in Medicine* **47**: 1083–1099.

Freedman, D. A. (1981). Bootstrapping regression models, *Annals of Statistics* **9**: 1218–1228.

Gençay, R., Selçuk, F. and Whitcher, B. (2002). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press London.

George, A. and Liu, J. (1981). *Computer Solution of Large Sparse Positive Definite Systems*, Prentice Hall Series in Computational Mathematics, Prentice-Hall, Englewood Cliffs, New Jersey.

Gössl, C., Fahrmeir, L., Pütz, B., Auer, L. M. and Auer, D. P. (2002). Fiber tracking from DTI using linear state space models: Detectability of the pyramidal tract, *NeuroImage* **16**: 378–388.

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*, 3rd edn, Oxford University Press.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Gudbjartsson, H. and Patz, S. (1995). The rician distribution of noisy MRI data, *Magnetic Resonance in Medicine* **34**: 910–914.

Hagmann, P., Thiran, J., Jonasson, L., Vandergheynst, P., Clarke, S., Maeder, P. and Meuli, R. (2003). DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection, *NeuroImage* **19**: 545–554.

Hahn, K., Prigarin, S., Heim, S. and Hasan, K. (2006). Random noise in diffusion tensor imaging, its destructive impact and some corrections, *in* J. Weickert and H. Hagen (eds), *Visualization and Processing of Tensor Fields*, Mathematics and Visualization, Springer-Verlag, New York, pp. 107–111.

Hahn, K., Prigarin, S. and Pütz, B. (2001). Edge preserving regularization and tracking for diffusion tensor imaging, *in* W. J. Niessen and M. A. Viergever (eds), *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'01)*, Utrecht, The Netherlands, pp. 195–203.

Hämmerlin, G. and Hoffmann, K. (1989). *Numerische Mathematik*, Springer-Verlag, New York.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*, Vol. 129 of *Lecture Notes in Statistics*, Springer-Verlag, New York.
URL: *http://www.xplore-stat.de/ebooks/scripts/wav/html/index.html*

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Monographs on Statistics and Applied Probability, 1st edn, Chapman and Hall, London.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models (with discussion), *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **55**: 757–796.

Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer-Verlag, New York.

Heim, S., Hahn, K., Auer, D. P. and Fahrmeir, L. (2004). Optimization of fiber tracking in human brain mapping: Statistical challenges, *in* A. Biggeri, E. Dreassi, C. Lagazio

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

and M. Marchi (eds), *Statistical Modelling. Proceedings of the 19th International Workshop on Statistical Modelling (IWSM'04)*, Florence, Italy, pp. 296–300.

Horsfield, M. A. and Jones, D. K. (2002). Applications of diffusion-weighted and diffusion tensor MRI to white matter diseases - a review, *NMR in Biomedicine* **15**: 570–577.

Hosey, T., Williams, G. and Ansorge, R. (2005). Inference of multiple fiber orientations in high angular resolution diffusion imaging, *Magnetic Resonance in Medicine* **54**: 1480–1489.

Hutchinson, M. F. and de Hoog, F. R. (1985). Smoothing noisy data with spline functions, *Journal of Numerical Mathematics* **47**: 99–106.

Jansen, M. (2001). *Noise Reduction by Wavelet Thresholding*, Vol. 161 of *Lecture Notes in Statistics*, Springer-Verlag, New York.

Johnstone, I. M. and Silverman, B. W. (2005). Ebayesthresh: R programs for empirical bayes thresholding, *Journal of Statistical Software* **12**: 1–38.

Jones, D. K. (2003). Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI, *Magnetic Resonance in Medicine* **49**: 7–12.

Jones, D. K. and Pierpaoli, C. (2005). Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a boostrap approach, *Magnetic Resonance in Medicine* **53**: 1143–1149.

Koay, C. G., Carew, J. D., Alexander, A. L., Basser, P. J. and Meyerand, M. E. (2006). Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging, *Magnetic Resonance in Medicine* **55**: 930–936.

Lang, S., Fronk, E. M. and Fahrmeir, L. (2002). Function estimation with locally adaptive dynamic models, *Computational Statistics* **17**: 479–499.

Lazar, M. and Alexander, A. L. (2003). An error analysis of white matter tractography methods: synthetic diffusion tensor field simulations, *NeuroImage* **20**: 1140–1153.

Lazar, M. and Alexander, A. L. (2005). Bootstrap white matter tractography (BOOT-TRAC), *NeuroImage* **24**: 524–532.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Le Bihan, D. (2003). Looking into the functional architecture of the brain with diffusion MRI, *Nature reviews. Neuroscience* **4**: 469–480.

Leemans, A., Sijbers, J., Verhoye, M., Van der Linden, A. and Van Dyck, D. (2005). Mathematical framework for simulating diffusion tensor MR neural fiber bundles, *Magnetic Resonance in Medicine* **53**: 944–953.

Lori, N. F., Akbudak, E., Shimony, J. S., Cull, T. S., Snyder, A. Z., Guillory, R. K. and Conturo, T. E. (2002). Diffusion tensor fiber tracking of human brain connectivity: Acquisition methods, reliability analysis and biological results, *NMR in Biomedicine* **15**: 493–515.

Lu, Y., Aldroubi, A., Gore, J. C., Anderson, A. W. and Ding, Z. (2006). Improved fiber tractography with bayesian tensor regularization, *NeuroImage* **31**: 1061–1074.

Mangin, J. F., Poupon, C., Clark, C., Le Bihan, D. and Bloch, I. (2002). Distortion correction and robust tensor estimation for MR diffusion imaging, *Medical Image Analysis* **6**: 191–198.

Marron, J. S. and Tsybakov, A. B. (1995). Visual error criteria for qualitative smoothing, *Journal of the American Statistical Association* **90**: 499–507.

Mechelli, A., Price, C. J., Friston, K. J. and Ashburner, J. (2005). Voxel-based morphometry of the human brain: Methods and applications, *Current Medical Imaging Reviews* **1**: 105–113.

Mishra, A., Yonggang, L., Jiingjing, M., Anderson, A. W. and Ding, Z. (2006). Unified framework for anisotropic interpolation and smoothing of diffusion tensor images, *NeuroImage* **31**: 1525–1535.

Mori, S. and Barker, P. B. (1999). Diffusion magnetic resonance imaging: Its principle and applications, *The Anatomical Record* **257**: 102–109.

Mori, S. and van Zijl, P. C. M. (2002). Fiber tracking: Principles and strategies - a technical review, *NMR in Biomedicine* **15**: 468–480.

Moseley, M. (2002). Diffusion tensor imaging and aging - a review, *NMR in Biomedicine* **15**: 553–560.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Ogden, R. T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhäuser Boston.

O'Sullivan, M., Jones, D. K., Summers, P. E., Morris, R. G., Williams, S. C. R. and Markus, H. S. (2001). Evidence for cortical "disconnection" as a mechanism of age-related cognitive decline, *Neurology* **57**: 632–638.

Pajevic, S., Aldroubi, A. and Basser, P. J. (2002). A continuous tensor field approximation of discrete DT-MRI data for extracting microstructural and architectural features of tissue, *Journal of Magnetic Resonance* **154**: 85–100.

Pajevic, S. and Basser, P. J. (2003). Parametric and nonparametric statistical analysis of DT-MRI data, *Journal of Magnetic Resonance* **161**: 1–14.

Papadakis, N. G., Martin, K. M., Wilkinson, I. D. and Huang, C. L. (2003). A measure of curve fitting error for noise filtering diffusion tensor MRI data, *Journal of Magnetic Resonance* **164**: 1–9.

Papadakis, N. G., Xing, D., Huang, C. L.-H., Hall, L. and Carpenter, T. A. (1999). A comparative study of acquisition schemes for diffusion tensor imaging using MRI, *Journal of Magnetic Resonance* **137**: 67–82.

Parker, G. J. M., Schnabel, J. A., Symms, M. R., Werring, D. J. and Barker, G. J. (2000). Nonlinear smoothing for the reduction of systematic and random errors in diffusion tensor imaging, *Magnetic Resonance Imaging* **11**: 702–710.

Pierpaoli, C. and Basser, P. J. (1996). Toward a quantitative assessment of diffusion anisotropy, *Magnetic Resonance in Medicine* **36**: 893–906.

Pinheiro, J. C. and Bates, D. M. (1996). Unconstrained parametrizations for variance-covariance matrices, *Statistics and Computing* **6**: 289–296.

Poupon, C., Clark, C. A., Frouin, V., Regis, J., Bloch, I., Le Bihan, D. and Mangin, J. (2000). Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles, *NeuroImage* **12**: 184–195.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Poupon, C., Mangin, J., Clark, C. A., Frouin, V., Regis, J., Le Bihan, D. and Bloch, I. (2001). Towards inference of human brain connectivity from MR diffusion tensor data, *Medical Image Anaylsis* **5**: 1–15.

Pütz, B., Ates, K. and Auer, D. P. (2003). TRAX — A program package for *in vivo* visualization of human brain fibers, *in* K. Kremer and V. Macho (eds), *Forschung und wissenschaftliches Rechnen: Beiträge zum Heinz-Billing-Preis 2002*, Vol. 62, Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen, pp. 75–85.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org*

Ruppert, D. and Carroll, R. J. (2000). Spatially-adaptive penalties for spline fitting, *Australian and New Zealand Journal of Statistics* **42**: 205–223.

Samsonov, A. A. and Johnson, C. R. (2004). Noise-adaptive nonlinear diffusion filtering of MR images with spatially varying noise levels, *Magnetic Resonance in Medicine* **52**: 798–806.

Schonberg, T., Pianka, P., Hendler, T., Pasternak, O. and Assaf, Y. (2006). Characterization of displaced white matter by brain tumors using combined DTI and fMRI, *NeuroImage* **30**: 1100–1111.

Schwartzman, A., Dougherty, R. F. and Taylor, J. E. (2005). Cross-subject comparison of principal diffusion direction maps, *Magnetic Resonance in Medicine* **53**: 1423–1431.

Sinha, S. S. and Schunk, B. G. (1992). A two-stage algorithm for discontinuity-preserving surface reconstruction, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**: 36–55.

Skare, S., Hedehus, M., Moseley, M. and Li, T.-Q. (2000). Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI, *Journal of Magnetic Resonance* **147**: 340–352.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Skare, S., Li, T., Nordell, B. and Ingvar, M. (2000b). Noise considerations in the determination of diffusion tensor anisotropy, *Magnetic Resonance Imaging* **18**: 659–669.

Sotak, C. H. (2002). The role of diffusion tensor imaging in the evaluation of ischemic brain injury - a review, *NMR in Biomedicine* **15**: 561–569.

Stejskal, E. O. and Tanner, J. E. (1965). Spin diffusion measurements: Spin echoes in the presence of time-dependent field gradient, *The Journal of Chemical Physics* **42**: 288–292.

Su, W. and Ward, R. K. (2006). An edge-based image interpolation approach using symmetric biorthogonal wavelet transform, *Proceedings of the 8th IEEE International Workshop on Multimedia Signal Processing (MMSP'06)*, Victory, BC/Canada.

Sullivan, E. V., Adalsteinsson, E., Hedehus, M., Ju, C., Moseley, M., Lim, K. O. and Pfefferbaum, A. (2001). Equivalent disruption of regional white matter microstructure in ageing healthy men and women, *Neuroreport* **12**: 99–104.

Tench, C. R., Morgan, P. S., Blumhardt, L. D. and Constantinescu, C. (2002). Impoved white matter fiber tracking using stochastic labeling, *Magnetic Resonance in Medicine* **48**: 677–683.

Toga, A. W. and Mazziotta, J. C. (eds) (2002). *Brain Mapping: The Methods*, 2nd edn, Academic Press, San Diego.

Toledo, S., Rotkin, V. and Chen, D. (2003). *Taucs: A library of sparse linear solvers*, Tel-Aviv University, Israel.
URL: *http://www.tau.ac.il/ stoledo/taucs/*

Tschumperlé, D. and Deriche, R. (2003). Variational frameworks for DT-MRI estimation, regularization and visualization, *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV'03)*, Nice, France, pp. 116–121.

Tuch, D. S. (2004). Q-ball imaging, *Magnetic Resonance in Medicine* **52**: 1358–1372.

Van de Ville, D., Blu, T. and Unser, M. (2006). Surfing the brain—An overview of wavelet-based techniques for fMRI data analysis, *IEEE Engineering in Medicine and Biology Magazine* **25**: 65–78.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Virta, A., Barnett, A. and Pierpaoli, C. (1999). Visualizing and characterizing white matter fiber structure and architecture in the human pyramidal tract using diffusion tensor MRI, *Magnetic Resonance Imaging* **17**: 1121–1133.

Wang, Z., Vemuri, B. C., Chen, Y. and Mareci, T. H. (2004). A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI, *IEEE Transactions on Medical Imaging* **23**: 930–939.

Weaver, J. B., Yansun, X., Healy, D. M. and Cromwell, L. D. (1991). Filtering noise from images with wavelet transforms, *Magnetic Resonance in Medicine* **21**: 288–295.

Weickert, J. and Hagen, H. (eds) (2006). *Visualization and Processing of Tensor Fields*, Mathematics and Visualization, Springer-Verlag, New York.

Westerhausen, R., Walter, C., Kreuder, F., Wittling, R. A., Schweiger, E. and Wittling, W. (2003). The influence of handedness and gender on the microstructure of the human corpus callosum: A diffusion-tensor magnetic resonance imaging study, *Neuroscience Letters* **351**: 99–102.

Westin, C. F., Maier, S. E., Mamata, H., Nabavi, A., Jolesz, F. A. and Kikinis, R. (2002). Processing and visualization for diffusion tensor MRI, *Medical Image Analysis* **6**: 93–108.

Whitcher, B. (2005). *waveslim: Basic wavelet routines for one-, two- and three-dimensional signal processing.* R package version 1.5.
URL: *http://www.image.ucar.edu/staff/whitcher/*
http://www.image.ucar.edu/staff/whitcher/book/

Winkler, G., Aurich, V., Hahn, K., Martin, A. and Rodenacker, K. (1999). Noise reduction in images: Some recent edge preserving methods, *Journal of Pattern Recognition and Image Analysis* **9**: 749–766.

Wirestam, R., Bibic, A., Lätt, J., Brockstedt, S. and Ståhlberg, F. (2006). Denoising of complex MRI data by wavelet-domain filtering: Application to high-b-value diffusion-weighted imaging, *Magnetic Resonance in Medicine* **56**: 1114–1120.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Curriculum Vitae

Name:	Susanne Heim
Geburtstag:	20. April 1976
Geburtsort:	Rosenheim

Schulbildung

1982 – 1986	Grundschule in Zellingen
1986 – 1995	Siebold-Gymnasium Würzburg, Neusprachlicher Zweig

Studium

1995 – 1997	Studium der Kulturpädagogik an der Universität Hildesheim *Hauptfach:* Bildende Kunst; *Abschluss:* Vordiplom
1997 – 2002	Studium der Statistik an der Universität Dortmund *Nebenfach:* Theoretische Medizin; *Abschluss:* Diplom
1999 – 2000	Auslandssemester an der University of Sheffield, UK
2000 – 2002	Studentische Hilfskraft bei Prof. Dr. J. Hartung und Prof. Dr. W. Urfer

Beruf

2002 – 2004	Mitarbeiterin der Arbeitsgruppe Kernspintomographie unter Leitung von PD Dr. D. P. Auer am Max-Planck-Institut für Psychatrie, München
2002 – 2006	Wissenschaftliche Mitarbeiterin im Sonderforschungsbereich 386 bei Prof. Dr. L. Fahrmeir an der Ludwig-Maximilians-Universität München
Seit 2007	Wissenschaftliche Mitarbeiterin am Institut für Statistik bei Prof. Dr. L. Fahrmeir an der Ludwig-Maximilians-Universität München

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.