Faculteit Wetenschappen
Departement Fysica

Optimal estimation of diffusion MRI parameters

Optimaal schatten van diffusie MRI parameters

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen: Fysica
aan de Universiteit Antwerpen te verdedigen door
Jelle VERAART

Promotor:
Prof. Dr. Jan Sijbers

Antwerpen

Doctoral committee:

Chairman:
Prof. Dr. Jacques Tempere

Members:
Prof. Dr. Jan Sijbers
Prof. Dr. Marleen Verhoye

External members:
Prof. Dr. Santiago Aja-Fernández
Prof. Dr. Jens H. Jensen
Prof. Dr. Stefan Sunaert

Contact information:

✉ Jelle Veraart
iMinds Vision Lab, Dept. of Physics
University of Antwerp
Universiteitsplein 1, N 1.16
2610 Wilrijk, Belgium
Jelle.Veraart@ua.ac.be
☎ +32 (0) 3 265 24 77
+32 (0) 3 265 22 45
http://www.visielab.ua.ac.be/people/jelle-veraart

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, broadcasting or by any other information storage and retrieval system without written permission from the copyright owner.

Cover illustration:
Quantitative distortion of diffusion MR parameters due to Rician noise

Dankwoord

Op basis van mijn eigen ervaring denk ik te kunnen stellen dat een doctoraatsonderzoek enkel succesvol afgerond kan worden indien je omringd bent door de juiste mensen. Enerzijds heb je mensen nodig die op basis van eerlijke en gefundeerde kritiek je wetenschappelijke ideëen mee helpen vormgeven. Anderzijds is er nood aan een groep mensen die je voortdurend blijven motiveren en steunen, in het bijzonder op één van die vele momenten dat de moed en enthousiasme om het werk af te ronden ver zoek zijn. In het algemeen zou je kunnen vermoeden dat die eerste groep mensen ook wel eens de collega's, zowel binnen als buiten de muren van het Visielab, genoemd kunnen worden, terwijl de steun en motivatie op de schouders van vrienden en familie terecht komen. Wel, ik kan na de voorbije jaren met heel veel plezier stellen dat het onderscheid tussen collega's en vrienden nog moeilijk te maken is. Bedankt aan alle mensen waarmee ik de voorbije jaren in contact ben gekomen, jullie hebben allemaal – direct, dan wel indirect – bijgedragen tot dit werk. Enkele mensen dienen in het bijzonder bedankt te worden:

In de eerste plaats wens ik Jan te bedanken. Bijna vijf jaar geleden was hij degene die me de kans heeft gegeven om bij het Visielab te starten. Hij gaf me als promotor de ideëen en het kader om mijn onderzoek aan te vatten. Zijn onuitputtelijke werklust en enthousiasme maakte hem continu bereikbaar voor feedback en ondersteuning om dit werk te maken tot wat het vandaag is.

Zonder de hulp van mijn dMRI collega's had het afronden van dit doctoraatonderzoek nog wel enkele jaren langer in beslag genomen. Bedankt, Dirk en Wim! Jullie voortdurende bereidheid om jullie kennis, enthousiasme voor dMRI en Matlab code te delen maakte een relatief vlotte start van mijn onderzoek mogelijk. Ondertussen heb ik niet meer het geluk om dagdagelijks met jullie te mogen samenwerken, maar het blijft een plezier jullie tegen te komen op conferenties. Een speciale vermelding voor Ben is op zijn plaats. Ben, vanaf mijn eerste werkdag deelde ik een bureau met jou. Het was een wetenschappelijk en vriendschappelijk plezier om met jou te kunnen samenwerken. Niet in het minst omdat het plezier tot ver buiten de muren van het Visielab reikte. Bedankt voor de gezellige dagen en avonden in de vele uithoeken die deze wereld rijk is.

Dit werk is grotendeels mee vormgegeven door interessante samenwerkingen en discussies met externe collega's. Ik zou het in het bijzonder Alexander, Marleen, Sofie en Stefan willen bedanken. Ik ben erg optogen over het feit dat onze samenwerkingen tot enkele gemeenschappelijke publicaties hebben mogen leiden. Alexander, ik zou je verder willen bedanken voor onze vele gesprekken op conferenties and Skype. Een weg zoeken in de academische wereld werd toch iets eenvoudiger dankzij jouw

goede raad, verfrissende inzichten en continu opbouwende kritiek.

Ik dank ook het IWT (Instituut voor de Aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen) voor de verleende financiële steun.

Een minder onderzoeksgericht luik van dit dankwoord begint met een welgemeende *dank u* richting alle collega's en oud-collega's van het Visielab. Danzij jullie allen was het Visielab de voorbije jaren meer dan mijn werkplek. De babbels tijdens de koffiepauzes, de avondjes – of waren het nachtjes – uit en het recent ingevoerde maandelijkse voetbalwedstrijdje zorgde voor een zeer aangename balans tussen werk en ontspanning. Diezelfde balans vond ik op de vele conferenties en workshops dankzij Alexander, Ben, Cathy, Chantal, Gwendolyne, Hilde, Judith, Kathleen, Laurena, Louise, Maarten, Sabine, Sjoerd, Stefan, Thibo, Thijs, Timo, Tom, Quinten en Wim. Het was een plezier om met jullie de conferentiedagen te kunnen aflsuiten op etentjes en feestjes.

Het meeste energie haalde ik wellicht uit de vele reizen, concerten, festivals, toogmomenten en sportieve uitjes met mijn vrienden. Jullie weten wie jullie zijn en wat jullie voor mij betekenen! Bedankt dat ik steeds op jullie kan rekenen. Graag zou ik ook mijn ouders bedanken voor de kansen en vrijheid die ze me voortdurend gaven om mijn weg te zoeken en dromen na te streven. Pa, Ma, het doet me deugd te weten dat er een warme thuis is waar ik op elk moment terecht kan, voor welke reden dan ook. Bedankt daarvoor!

De laatste woorden van dit dankwoord wil ik richten tot mijn vriendin ... al zou ik niet weten hoe ik mijn dank aan haar kan vervatten in woorden. Anneleen, bedankt voor alles! Bedankt voor het geduld dat je toonde in die momenten dat ik rusteloos mezelf aan het zoeken was, bedankt voor de onvoorwaardelijke steun en motivatie, bedankt voor de thuis die je van ons huis gemaakt hebt de voorbije jaren. Ik besef maar al te goed dat ik zonder jou niet de persoon zou zijn die ik nu ben en daarvoor ben ik je enorm dankbaar.

Bedankt!
Jelle

Contents

Summary

Diffusion magnetic resonance imaging (dMRI) is currently the method of choice for the in vivo and non-invasive quantification of water diffusion in biological tissue. Several diffusion models have been proposed to obtain quantitative diffusion parameters, which have shown to provide novel information on the structural and organizational features of biological tissue, the brain white matter in particular. The goal of this dissertation is to improve the **accuracy** of the diffusion parameter estimation, given the non-Gaussian nature of the diffusion-weighted MR data. In part I of this manuscript, the necessary basics of dMRI are provided. Next, Part II deals with diffusion parameter estimation and includes the main contributions of the research. Finally, Part III covers the construction of a population-based dMRI atlas of the rat brain.

Diffusion MRI: the basics

Chapter 1 briefly introduces magnetic resonance imaging (MRI), the medical imaging technique that utilizes a strong static magnetic field and radio frequent electromagnetic waves to create images based on local properties of water molecules and their mutual interactions.

One property of interest is the local mobility of water molecules. The sensitization of a MR image to the *molecular diffusion* is called diffusion-weighted MRI (or diffusion MRI, dMRI). Statistically spoken, the signal intensity of each voxel in a diffusion MR image depends on the displacement probability distribution function (PDF) of water molecules captured in the imaged volume. The knowledge of that displacement PDF would provide information about the geometry of the underlying tissue microstructure, at scales much smaller than the imaging resolution. For example, in fibrous tissue, water molecules tend to diffuse more along the fibers, enabling researchers to obtain information about the orientation and *integrity* of the underlying tissue. Unfortunately, due to hardware limitations and time constraints, the full displacement PDF can generally not be computed. However, the cumulant expansion of the diffusion-weighted signal allows extracting several statistics from that PDF. Indeed, the second order cumulant expansion – cf. Diffusion Tensor Imaging (DTI) – provides the standard deviation of the displacement, which relates to the diffusion coefficient by the Einstein equation. Extending the series expansion to the fourth order – cf. Diffusion Kurtosis Imaging (DKI) – would reveal the kurtosis of the displacement PDF. Both diffusion and kurtosis values have shown to be useful in the diagnosis of stroke and to investigate white matter pathologies such as brain tumors. The reader is referred to Chapter 2 for a detailed review of the principles of diffusion MRI and diffusion modeling.

The quantitative diffusion measures have the highest clinical value – especially in multi-center clinical trials – if they can be extracted precisely and accurately.

To date, that is challenging due to the low signal-to-noise ratio (SNR) of the diffusion-weighted MR data. Indeed, at low SNR, the measured MR signals cannot be assumed to be normally distributed variables. Although the actual data distribution is often assumed to be Rician, it actually depends on the applied image formation technique – e.g. single v. multichannel imaging or image v. frequency space reconstruction. The possible data distributions are discussed in Chapter 3.

Diffusion MRI: parameter estimation

The diffusion models, described in Chapter 2, are mathematical relations that describe the diffusion-weighted MR images in terms of diffusion parameters. Those parameters need to be estimated from a series of acquired diffusion-weighted MR images. However, acquired MR signals are disturbed by noise. In Chapter 4, a theoretical introduction to estimators, i.e. methods to extract information about model parameters from noisy measurements, is given. In Chapter 5, the strengths and limitations of the popular class of least squares estimators in context of dMRI is discussed. Special attention goes to the weighted linear least squares estimator. Indeed, the DTI and DKI model have in common that they can be structured into a linear regression form depending on the natural logarithm of the diffusion-weighted MR signals. The (weighted) linear least squares estimator has a closed-form solution and a low computation cost. Moreover, the linear estimators are very accurate, especially compared to the nonlinear alternative if the SNR of all (Rician distributed) data exceeds two. If that condition is not met, more advanced estimators need to be considered. A typical example is the maximum likelihood estimator (MLE). The MLE has some desirable properties such as consistency, asymptotic efficiency and asymptotic normality. Those properties stem from its basis on the joint PDF of the diffusion-weighted data. Unfortunately, the necessity of data correction (e.g. motion and eddy current corrections) prior to model fitting causes the MLE's dependency on the joint PDF to become a weakness because the altered data PDF can no longer be expressed analytically. A practical alternative to the MLE, i.e. the conditional least squares estimator, is introduced in Chapter 6. These *advanced* diffusion parameter estimators require the knowledge of the noise parameter. The estimation of the noise parameter has become very challenging due to the use of parallel imaging techniques. Indeed, the noise parameter is generally spatially varying. As such, a 3D noise map must be estimated from the diffusion-weighted images. The development of such 3D noise map estimation strategy is described in Chapter 7. In Chapter 8, necessary constraints on the DKI model parameters are discussed.

Atlas construction

In this final part, an anatomically labeled DTI atlas of the adult rat brain is proposed. The atlas is constructed using a population based atlas approach to create a template, which represents the average anatomy. During the construction, a non-rigid coregistration technique is used to avoid local misalignment inaccuracies due to intersubject differences. The delineation of brain structures was performed on high resolution ex vivo scans and the resulting parcellation maps were non-linearly warped into the in-vivo atlas space afterwards.

Samenvatting

Diffusie magnetische resonantie beeldvorming (dMRI) is een unieke techniek die toelaat om de willekeurige bewegingen (of diffusie) van de waterstofkernen in biologische weefsels – vaak de witte hersenmaterie – niet-invasief te meten. Er werden reeds verschillende modellen geïntroduceerd om dit diffusieproces te kwantificeren met verschillende diffusie parameters. Op basis van deze parameters kan op een indirecte manier informatie verkregen worden over de geometrie van de onderliggende microstructuur. Het doel van deze thesis is het verbeteren van de **juistheid** van de diffusie parameterschatting, gegeven dat de diffusie-gewogen MR data niet normaal verdeeld zijn.

Diffusie MRI: de basis

In Hoofdstuk 1 worden kort de basisprincipes van MRI geïntroduceerd. MRI is een medische beeldvormingstechniek die gebruik maakt van een krachtige magneet en radiogolven om lokale eigenschappen van waterstofkernen en hun onderlinge interacties in kaart te brengen. Een interessante eigenschap is de lokale beweeglijkheid/diffusie, die indirect gemeten kan worden met dMRI. Het gemeten diffusie MR signaal is afhankelijk van de statistische waarschijnlijkheidsdistributie (PDF) van de verplaatsing van de waterstofkernen. Deze statistische verdeling bevat biologisch en klinisch relevante informatie omtrent de microstructuur van weefsels. Bijvoorbeeld, in weefsels met een sterke vezelstructuur zullen de waterstofkernen meer bewegen langsheen de vezels dan loodrecht daarop. Op basis van dit principe kan informatie verkregen worden over de oriëntatie en integriteit van de onderliggende microstructuur. Door hardware - en tijdsbeperkingen is het meestal niet mogelijk om de volledige PDF te berekenen. Echter, de cumulant reeksontwikkeling van het diffusie-gewogen signaal laat het toe om enkele eigenschappen van de PDF te bepalen. Zo zal de tweede-orde cumulantontwikkeling – cfr. diffusie tensor beeldvorming (DTI) – resulteren in de standaardafwijking van de verplaatsing. Deze standaardafwijking is rechtstreeks gekoppeld aan de diffusiecoefficiënt door de Einstein vergelijking. De vierde-orde cumulantonwikkeling – cfr. diffusie kurtosis beeldvorming (DKI) – levert daarenboven ook nog de kurtosis van de diffusie PDF op. Zowel de diffusie- als kurtosiscoefficiënten zijn interessante maten voor de diagnose van beroertes en voor het bestuderen van hersenaandoeningen zoals tumoren. Een gedetailleerde beschrijving van diffusie MRI en diffusie modelering is gegeven in Hoofdstuk 2. De kwantitatieve parameters hebben de hoogste klinische waarde indien ze een met een hoge juistheid en precisie bepaald kunnen worden. Dit is tot op vandaag niet vanzelfsprekend doordat diffusie-gewogen MR beelden inherent een lage signaal-ruisverhouding (SNR) hebben. Gemeten MR signalen met een lage SNR kunnen niet beschouwd worden als normaal verdeelde variabelen. De feitelijke verdeling van MR signalen hangt namelijk af van het opnameproces.

De mogelijke dataverdelingen worden besproken in Hoofdstuk 3.

Diffusie MRI: parameterschatting

De diffusiemodellen, geïntroduceerd in Hoofdstuk 2, leggen een wiskundige relatie tussen het verwachte diffusie-gewogen signaal en tal van diffusieparameters. In de praktijk dienen deze parameters geschat te worden op basis van een reeks gemeten diffusie-gewogen signalen. Deze signalen zijn echter verstoord door ruis. Een theoretische inleiding over parameterschatters – i.e. functies om parameters te bepalen op basis van ruizige data – is gegeven in hoofdstuk 4. In Hoofdstuk 5 wordt er dieper ingegaan op de kracht en beperkingen van de populaire klasse van de kleinste kwadratenschatters in het kader van dMRI. Speciale aandacht gaat naar de gewogen lineaire kleinste kwadratenschatter aangezien zowel DTI als DKI gelineariseerd kunnen worden door een log-transformatie. Lineaire kleinste kwadratenschatters hebben steeds een gesloten uitdrukking voor de oplossing en een lage computationele kost. Daarenboven zijn deze schatters mogelijks juister dan hun niet-lineaire variant. De SNR moet in dat geval wel hoger zijn dan 2. Indien niet aan deze voorwaarde voldaan kan worden, schakelt men best over naar meer geavanceerde schatters. Een typisch voorbeeld is de maximale waarschijnlijksheid-schatter (MLE). De MLE heeft optimale theoretische eigenschappen betreffende precisie en juistheid indien de verdeling van de data gekend is. Deze voorwaarde is in praktijk echter niet steeds vervuld aangezien parameterschatting vaak wordt voorafgegaan door bewegingscorrectie. Dergelijke beeldverwerkingsstappen zullen de dataverdeling vaak zodanig wijzigen dat ze niet meer analytisch uit te drukken zijn. Een praktisch alternatief voor MLE – de conditionele kleinste kwadratenschatter – wordt geïntroduceerd in Hoofdstuk 6. De geavanceerde schatters vereisen de kennis van het ruisniveau. De opmars van parallele beeldvormingstechnieken maakt het schatten van het ruisniveau extra uitdagend. Inderdaad, het ruisniveau is meestal spatiaal variërend waardoor een 3D ruismap geschat zal moeten worden op basis van de diffusie-gewogen data. Een nieuwe techniek om dit te verwezenlijken wordt voorgesteld in Hoofdstuk 7. In Hoofdstuk 8, zullen er noodzakelijke randvoorwaarden aan de DKI model paramterschatting besproken worden. Deze voorwaarden moeten garanderen dat de uitkomsten voldoen aan biologische en fysische randvoorwaarden.

Atlasconstructie

In dit laatste gedeelte wordt een anatomisch gelabelde DTI atlas van de hersenen van de Sprague Dawley rat voorgesteld. De atlas representeert de gemiddelde anatomie van de rathersenen en is bepaald op basis van DTI beelden van een gehele populatie. Anatomische verschillen tussen verschillende subjecten werden geminimaliseerd door gebruik te maken van niet-lineaire beeldregistratietechnieken. Hoge resolutie ex vivo diffusiegewogen beelden werden gebruikt voor de manuele aflijning van de anatomische structuren.

Part I

Diffusion MRI: the basics

CHAPTER 1

Magnetic resonance imaging

Contents

1.1 Introduction

Magnetic resonance imaging (MRI) is a powerful medical imaging technique that enables the visualization of structure and function in vivo because of large contrast within soft tissue without the need for ionizing radiation. A basic knowledge of MRI is required for understanding the concepts discussed in this thesis. Therefore, this introductory chapter briefly explains the principles of nuclear magnetic resonance (NMR) and image formation. Also, the basics of pulse sequence design and parallel imaging are shortly discussed here. For more in-depth information on these topics we refer to literature [e.g. Liang and Lauterbur, 2000].

1.2 History

The extraordinary journey of MRI started in the 1930s by pioneering work of **Isodor Isaac Rabi**, who was awarded by the Nobel Prize in Physics in 1944 for his observation of the nuclear magnetic resonance (NMR) phenomena in an artificial environment [Rabi et al., 1938]. This American physicist showed that one can manipulate and identify atomic nuclei, which behave like spinning tops whose orientation axes are aligned with an externally magnetic field, by exposing them to radio-waves. Almost a decade later, **Felix Bloch** and **Edward Mills Purcell** independently demonstrated that any solid or liquid can be placed in a magnetic field to identify the specific atoms without affecting it in any perceptible way using the NMR phenomena [Bloch et al., 1946, Purcell et al., 1946]. They were jointly awarded by the Nobel Prize in Physics in 1952. A promising technique was born. However, only at the seventies NMR signals could be used to generate two-dimensional (2D) images. **Paul Lauterbur** expanded upon the work of **Herman Carr** to develop spatial information encoding principles [Carr and Purcell, 1954, Lauterbur, 1973]. **Peter Mansfield** developed a method, currently known as echo planar imaging (EPI) to acquire such 2D images in only a few seconds [Mansfield, 1977]. Both scientist received the Nobel Prize in Physiology and Medicine in 2003 for their seminal contributions, which *led to the applications of magnetic resonance in medical imaging.* However, the Nobel Prize ensued some controversy [Dreizen, 2004, Macchia et al., 2007]. Why was **Raymond Damadian** not honored by the Nobel Prize for his contribution to MRI in medicine? The Armenian-American medical doctor showed differences in NMR properties among normal tissues and between normal and cancer tissues [Damadian, 1971]. Moreover, he was the first to achieve human whole-body MR images (granted a patent in 1974). Finally, in this brief overview of the main pioneers of NMR/MRI, one cannot disregard **Richard Ernst**, which was awarded by the Nobel Prize in Chemistry in 1991. In 1975 he described the use of Fourier transform to reconstruct 2D images, using switched magnetic field gradients in the time domain for spatially encoding [Kumar et al., 1975]. A more extensive overview of the history of MRI is given in *The pioneers of NMR and Magnetic Resonance in Medicine. The Story of MRI* by Mattson and Simon [1996].

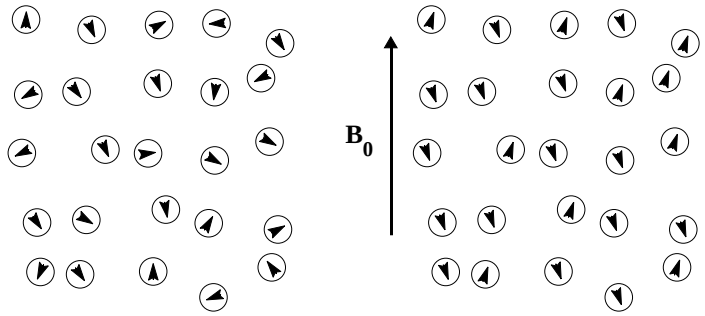


Figure 1.1: (a) Due to thermal motion, the magnetic dipoles moment will have random orientation. (b) When placed in a static magnetic field $\boldsymbol{B}_0$, the magnetic dipoles will align with the direction of $\boldsymbol{B}_0$.

1.3 Signal generation and detection

1.3.1 Spin physics

All atomic nuclei consisting of an odd number of protons or neutrons possess a spin angular momentum $\boldsymbol{J}$, often called *nuclear spin*. The nuclear spin is fundamental to the existence of a *magnetic dipole moment* $\boldsymbol{\mu}$. Both properties are linked by a nucleus-dependent constant, the gyromagnetic ratio γ:

$$\boldsymbol{\mu} = \gamma \boldsymbol{J}. \tag{1.1}$$

The magnitude of the magnetic dipole moment depends on the spin quantum number I, which is another intrinsic property of the nucleus. To exhibit the property of magnetic resonance, the nucleus must have a non-zero value of I. This is, for example, the case for ^{19}F, ^{13}C, ^{23}Na, or ^{31}P. Although all these nuclei can be studied with NMR, the most common nucleus of interest in MR research is the hydrogen nucleus (^{1}H), because of its high natural abundance under the form H_2O in biological tissues. For ^{1}H, the gyromagnetic ratio is 2.675×10^8rad/s/T, whereas ^{1}H has half-integer spin, i.e. $I = 1/2$. The ensemble of protons present in the object form a spin system. Such a spin system has no net magnetization in the absence of an external magnetic field. Indeed, due to thermal motion, the magnetic dipoles moment will have random orientation. When placed in a static magnetic field $\boldsymbol{B}_0$[1], the magnetic dipoles will align with the direction of $\boldsymbol{B}_0$, i.e. the z-axis. However, the magnetic dipole moments can take two possible orientations with respect to the z-axis: parallel and antiparallel (see Fig. 1.1). As a result, the z-component of the magnetic dipole can have two values:

$$\mu_z = \pm \frac{1}{2} \gamma \hbar, \tag{1.2}$$

[1]In a clinical setting, the strength of the magnetic field $\boldsymbol{B}_0$ is typically 1.5T or 3T.

with $\hbar$ Planck's constant divided by 2π. Next, driven by the external magnetic field, the magnetic dipole moments will precess about the z-axis. The angular frequency of the precession is better known as the Larmor frequency and given by:

$$\omega_0 = \gamma B_0. \tag{1.3}$$

In equilibrium, the phase of the rotation is random. Furthermore, it has been observed that spins that occupy a different state will show a different potential energy:

$$E = \boldsymbol{\mu} \cdot \boldsymbol{B}_0 = \begin{cases} E_\uparrow = -\frac{1}{2}\gamma\hbar B_0 \\ E_\downarrow = \frac{1}{2}\gamma\hbar B_0 \end{cases} \tag{1.4}$$

The nonzero difference in energy level – $\Delta E = \gamma\hbar B_0$ – is known as the *Zeeman splitting*. The two energy levels are commonly referred to as spin-up and spin-down with the spin-down state having higher energy than the spin-up state. Consequently, the spin-up state has higher prevalence:

$$\frac{\mathrm{N}_\uparrow}{\mathrm{N}_\downarrow} = \exp\frac{\Delta E}{k_B T} > 1, \tag{1.5}$$

with T the temperature and k_B the Boltzmann constant. Furthermore, $\mathrm{N}_\uparrow$ and $\mathrm{N}_\downarrow$ are the number of spins in the low and high-energy state, respectively. The difference in occupation of both states:

$$\mathrm{N}_\uparrow - \mathrm{N}_\downarrow \approx N_s \frac{\gamma\hbar B_0}{2k_B T}, \tag{1.6}$$

with N_s the total number of spins, is very small. Nevertheless, it is sufficient to generate an observable macroscopic magnetization vector $\boldsymbol{M}$. Being able to treat the behavior of all spins in the system in terms of a net magnetization vector $\boldsymbol{M}$ allows a classical description of NMR.

1.3.2 Bulk magnetization: a classical NMR description

The magnetization vector can be decomposed in an x, y, and z-component: $\boldsymbol{M} = [M_x, M_y, M_z]$. At equilibrium, both transverse components ($M_x(0)$ and $M_y(0)$) are zero because of the random phase of the individual magnetic dipole moments. The z-component, on the other hand, is nonzero:

$$M_z(0) = \frac{1}{2}\left(\mathrm{N}_\uparrow - \mathrm{N}_\downarrow\right)\gamma\hbar. \tag{1.7}$$

A magnetization vector, placed in an external magnetic field, is subject to a torque. Given that $\boldsymbol{B}_0$ is a static magnetic field along the z-axis, the resulting rate with which $\boldsymbol{M}$ changes in time is given by:

$$\frac{d\boldsymbol{M}}{dt} = \gamma\boldsymbol{M} \times \boldsymbol{B}_0 = \left[\omega_0 M_y, \omega_0 M_x, 0\right], \tag{1.8}$$

with ω_0 again the Larmor frequency [Bloch et al., 1946]. The solution of Eq. (1.8) gives an expression for the components of $\boldsymbol{M}(t)$:

$$\begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega_0 t) & \sin(\omega_0 t) & 0 \\ -\sin(\omega_0 t) & \cos(\omega_0 t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} M_x(0) \\ M_y(0) \\ M_z(0) \end{bmatrix} \tag{1.9}$$

If an electrically conducting coil is placed around the subject, perpendicular to the transverse plane, a rotating transverse magnetization component will induce a voltage in the coil whose amplitude is proportional to the magnitude of the transverse component (cf. Faraday law). So, to generate measurable signals in such a receiver coil, the magnetization vector should be tilted into the transverse plane. This tilting is a consequence of exciting spins out of equilibrium state and getting them in-phase, and is achieved by shortly turning on an additional – oscillating – magnetic field $\boldsymbol{B}_1$: the radio frequency (RF) pulse [Rabi et al., 1938]. Note that $\boldsymbol{B}_1$ is much weaker than $\boldsymbol{B}_0$ as its strengths is *only* about 50 mT.

1.3.3 RF Excitation

A RF pulse circularly oscillating in the xy-plane at the same frequency as the precessing spins carries the required amount of energy, that is ΔE, to induce a coherent transition of those spins from one energy state to another. The transition will change both the transversal and longitudinal components of the magnetization vector. The application of a rotating magnetic field $\boldsymbol{B}_1$ in distinction to the static $\boldsymbol{B}_0$ will change the motion of $\boldsymbol{M}$. Indeed, $\boldsymbol{M}$ will precess around $\boldsymbol{B}_1$ at an angular frequency $\omega_1 = \gamma B_1$, while $\boldsymbol{B}_1$ rotates at an angular frequency $\omega_0 = \gamma B_0$ about the z-axis. If a rectangular RF pulse is turned on during a period Δt, then $\boldsymbol{M}$ will be flipped in the tranverse plane by an angle $\alpha = \omega_1 \Delta t$. Often, the RF pulse is applied so that $\alpha = 90°$, i.e. the magnetization vector is perpendicular to the z-axis (see Fig. 1.2a). That pulse is called a 90°-pulse.

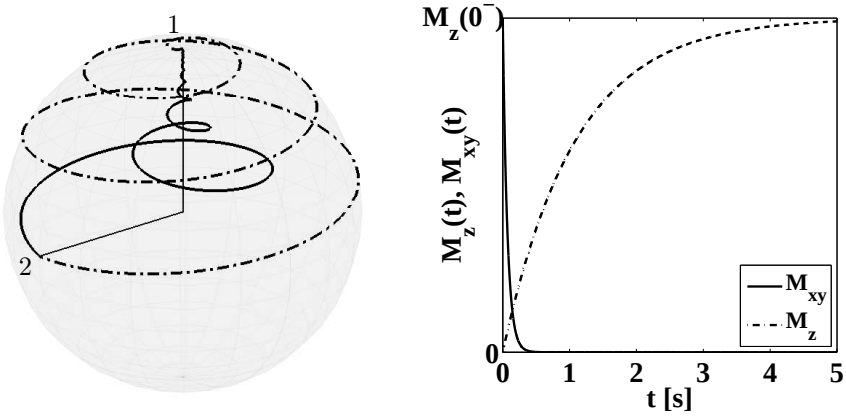
1.3.4 Relaxation

After turning off the RF pulse, the magnetized spin system will gradually return to the equilibrium state. This phenomena is called relaxation. Relaxation can be decomposed into the following processes.

Spin-lattice relaxation: An energy exchange between the spins and their surrounding environment by heat transfer;

Spin-spin relaxation: Loss of phase coherence of the magnetized spin system due to local field inhomogeneities, which are induced by surrounding spins.

The spin-lattice relaxation refers to the process in which the spins give the energy they obtained from the RF pulse back to the surrounding lattice, thereby restoring the equilibrium distribution of the populations of both spin states (cfr. Eq. (1.6)). The result is growth of M_z, characterized by the longitudinal relaxation time T_1. After applying a 90°-pulse, the longitudinal component of the magnetization vector



(a) (b)

Figure 1.2: (a) Trajectory followed by the tip of the magnetization vector $\boldsymbol{M}$ excitation (dashed line $1 \rightarrow 2$) and relaxation (solid line $2 \rightarrow 1$) given a 90°-pulse. (b) Transversal (solid line) and longitudinal (dashed line) relaxation after the 90°-pulse. Note that longitudinal relaxation is a much slower process than transversal relaxation. T_1 and T_2 relaxation times at 3T in the human white matter are approximately 1000 ms and 60 ms, respectively [Stanisz et al., 2005].

evolves in function of the time t:

$$M_z(t) = M_z(0^-)\left[1 - \exp\left(-\frac{t}{T_1}\right)\right], \tag{1.10}$$

with $M_z(0^-)$ the longitudinal magnetization at equilibrium. The spin-spin relaxation is the result of local fluctuations of the magnetic field induced by surrounding magnetized spins. Indeed, the magnetic moments of the spins are superimposed to the main magnetic field. Such local field fluctuations cause temporal variations in precession frequency that depahse the spins. The process finds expression in an exponential decay of the transversal component of the magnetization vector. The exponential decay is characterized by a time constant T_2. After applying an RF-pulse, the evolution of M_x and M_y in time is given by:

$$\begin{aligned} M_x(t) &= M_x(0^+)\sin(\omega_0 t)\exp\left(-\frac{t}{T_2}\right) \\ M_y(t) &= M_y(0^+)\cos(\omega_0 t)\exp\left(-\frac{t}{T_2}\right), \end{aligned} \tag{1.11}$$

with $M_x(0^+)$ and $M_y(0^+)$ the x and y-component of the magnetization vector immediately after turning off the RF pulse [Bloch et al., 1946]. In case of a 90°-pulse, the total transversal component $M_{xy}(0^+)$ equals $M_z(0^-)$. Note that this transversal component, $M_{xy}(t) = M_x(t) + iM_y(t)$, keeps precessing around the z-axis with a constant angular frequency equal to the original Larmor frequency

(ω_0), while its magnitude decreases exponentially (see Fig. 1.2b):

$$M_{xy}(t) = M_{xy}(0^+) \exp\left(-\frac{t}{T2}\right) \exp\left(-i\omega_0 t\right) \tag{1.12}$$

Due to heterogeneity of the sample and inhomogeneity of the magnetic field, the decay will be faster than T_2 though. The actual relaxation time is T_2^*. Anyhow, a voltage (signal) with exponentially decaying magnitude will be induced in an electrically conducting coil that is positioned in the transverse plane. This time-dependent signal is called free induction decay (FID).

1.3.5 Spin echo

Relaxation times vary across different healthy and pathological tissue types [Damadian, 1971]. Therefore, both T_1 and T_2 are besides the proton density main intrinsic factors to determine contrast in MR signals. A common way to generate T_2 -weighted contrast is the application of a pulse sequence that is well-known as the *spin-echo* sequence [Hahn, 1950]. The spin echo sequence is schematically presented in Fig. 1.3. After applying a 90°-pulse, the spins start to dephase due to all effects contributing to the T_2^* relaxation. At TE/2 – with TE short for *echo time* – the magnetization is flipped by applying a 180°-pulse. After another period of TE/2, the spins are rephasing, thus producing a measurable echo signal. The signal decay at TE, compared to the start of the experiment, now solely originates in the T_2-relaxation.

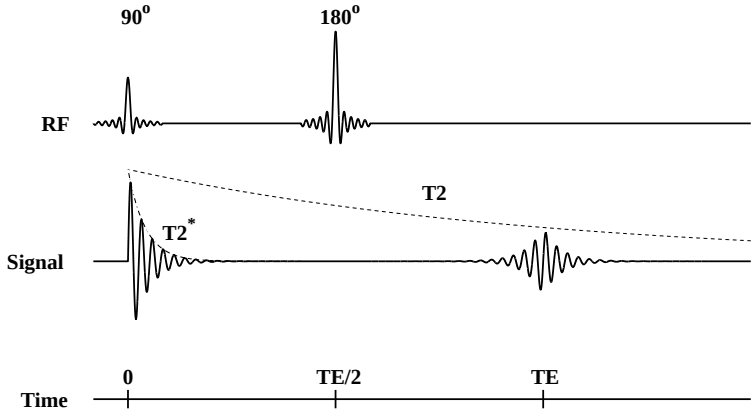


Figure 1.3: Spin echo sequence

1.4 MR image formation

1.4.1 Magnetic field gradients

Spatial localization of signals is essential to move from MR signals to MR images [Carr and Purcell, 1954, Lauterbur, 1973]. Two ways to encode spatial information will shortly be discussed:

Frequency encoding: The spins' precessing rate is made linearly dependent to (a component of) the spatial location by applying magnetic field gradients. Such magnetic field gradients are static magnetic fields whose strength varies linearly across a region of space. Let's consider a frequency encoding gradient G_x, applied along the x-axis. The Larmor frequency at position x is now given by:

$$\omega(x) = \omega_0 + \gamma G_x x \tag{1.13}$$

All spins in an infinitesimal neighborhood dx of x will generate following signal:

$$d\mathcal{K}(x,t) = \rho(x)dx \exp\left(-i\omega(x)t\right), \tag{1.14}$$

with time t and $\rho(x)$ the (weighted) spin-density at location x. Note that, for simplicity, the relaxation effects are ignored. The received signal for the entire object thus becomes:

$$\begin{aligned}
\mathcal{K}(t) &= \int_{object} d\mathcal{K}(x,t) \\
&= \int_{-\infty}^{\infty} \rho(x) \exp\left(-i\omega(x)t\right) dx \\
&= \underbrace{\exp\left(-i\omega_0 t\right)}_{\text{carrier signal}} \int_{-\infty}^{\infty} \rho(x) \exp\left(-i\gamma G_x x t\right) dx
\end{aligned} \tag{1.15}$$

Phase encoding: Alternatively, the signal can be spatially encoded by shortly applying a magnetic field gradient for a given time, τ, after the application of a RF pulse. Let's now assume a phase encoding gradient G_y was applied along the y-axis. After time τ, a phase offset to the magnetization vector, which depends on the y-position, has been introduced:

$$\phi(y) = \gamma G_y y \tau. \tag{1.16}$$

The received signal for the entire object is now given by:

$$\begin{aligned}
\mathcal{K}(t) &= \int_{-\infty}^{\infty} \rho(y) \exp\left(-i\left(\omega_0 t + \phi(y)\right)\right) dy \\
&= \underbrace{\exp\left(-i\omega_0 t\right)}_{\text{carrier signal}} \int_{-\infty}^{\infty} \rho(y) \exp\left(-i\phi(y)\right) dy
\end{aligned} \tag{1.17}$$

Quadrature detection and signal demodulation: The detection of complex-valued MR signals must be done with a *phase sensitive* receiver in order not to lose any encoded information [Hoult et al., 1984]. Such a receiver is called a quadrature detector and consists out of two antennae, which are positioned perpendicular to each other. One antenna receives the cosinusoidal component of the complex exponential signal function, the other the sinusoidal. In both components, the high frequency carrier signal is removed by applying a low-pass filter after mixing the signal with a cosine oscillating at the Larmor frequency. This is called signal demodulation. From this point on, the carrier signals in Eq. (1.15) and Eq. (1.17) are assumed to be crossed out.

1.4.2 k-space

Let's return to the (demodulated) signal given in Eq. (1.15). By substituting

$$k_x = \frac{\gamma}{2\pi} G_x t \tag{1.18}$$

into Eq. (1.15), the frequency encoded signal is given by:

$$\mathcal{K}(k_x) = \int_{-\infty}^{\infty} \rho(x) \exp\left(-i2\pi k_x x\right) dx. \tag{1.19}$$

Clearly, there is a Fourier relationship between $\rho(x)$ and the received signal in *k-space* [Kumar et al., 1975]. Similar conclusions can be drawn for the (demodulated) phase encoded signal. After substituting:

$$k_y = \frac{\gamma}{2\pi} G_y \tau, \tag{1.20}$$

the phase encoded signal can be written as

$$\mathcal{K}(k_y) = \int_{-\infty}^{\infty} \rho(y) \exp\left(-i2\pi k_y y\right) dy. \tag{1.21}$$

The application of frequency and phase encoding along the x and y-direction, allows two-dimensional (2D) imaging:

$$\mathcal{K}(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho(x, y) \exp\left(-i2\pi \left(k_x x + k_y y\right)\right) dx dy. \tag{1.22}$$

$\mathcal{K}(k_x, k_y)$ is often called the k-space. After sampling the k-space at several frequencies, a 2D image showing the local spin densities, can be computed by the inverse Fourier transformation $\mathcal{F}^{-1}$ [Kumar et al., 1975] (see Fig. 1.4(a,b)).

There are many possible schemes for sampling the k-space, each scheme having its own strengths and limitations. In this thesis, a Cartesian sampling scheme

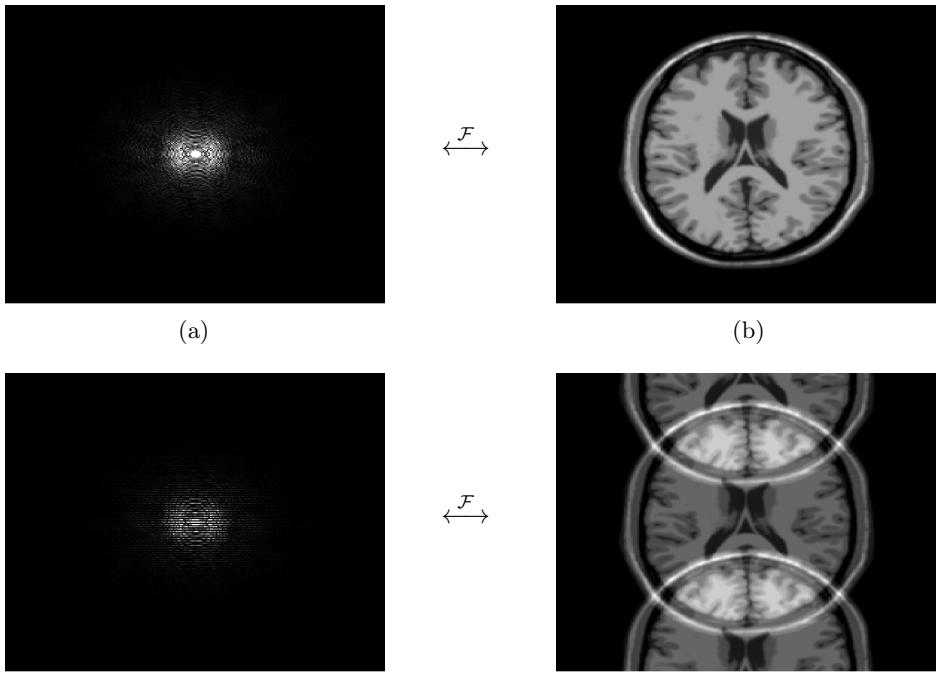


Figure 1.4: The MR image (a) can be computed as the inverse Fourier transform of a fully sampled k-space (b). Both the k-space and image are complex-values. However, only their magnitudes are shown. Undersampling of the k-space (c) will result in aliasing in the image space (d).

will always be applied. Specifically, each spin-echo signal is first phase-encoded along the y-axis by shortly switching on G_y. Next, the signal is acquired in the presence of a frequency-encoding gradient G_x. A fast implementation of the scheme is used in single-shot Echo Planar Imaging (ss-EPI) [Mansfield, 1977]. After a single excitation, the full k-space is efficiently traversed using time-varying gradients. Between positive and negative lobes of the frequency encoding gradient G_x, a short phase encoding *blip* is applied to step along the y-direction of the k-space (see Fig. 1.5). The trajectory of this so-called blipped EPI sequence is given in Fig. 1.6. ss-EPI is an acquisition technique prone to several imaging artifacts. Nevertheless, the use ss-EPI in diffusion MRI is encouraged by its low sensitivity to motion artifacts, which result in image blurring, signal drop-outs and ghosting. In the next chapter, the EPI artifacts that are important for diffusion MRI, will be discussed. A complete overview of all MRI artifacts goes beyond the scope of this work.

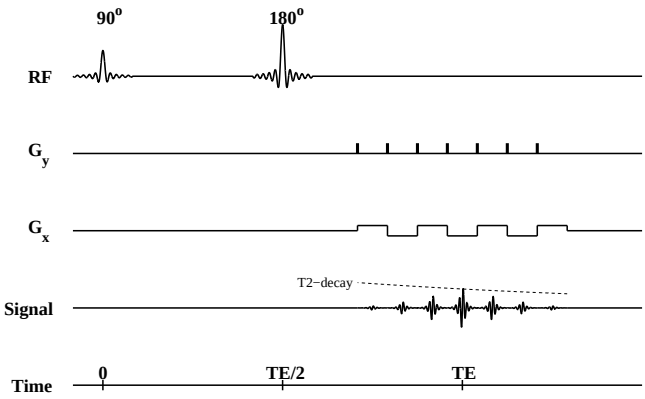
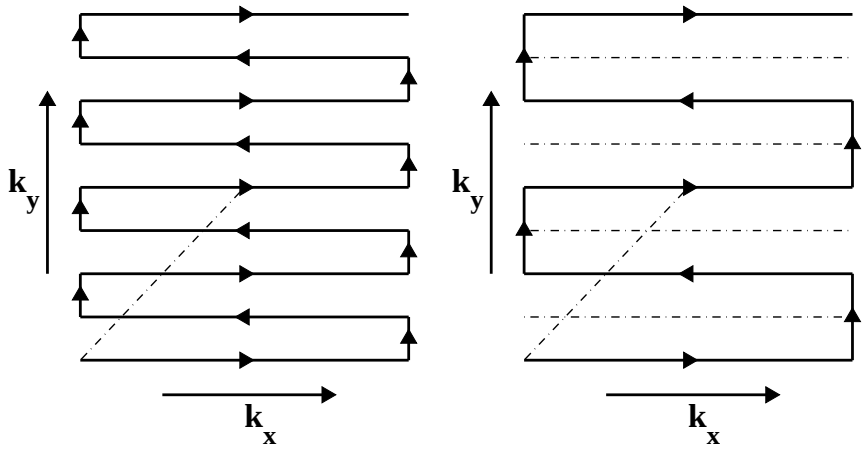


Figure 1.5: Blipped ss-EPI



(a) (b)

Figure 1.6: EPI k-space traversal with subsampling factor $R = 1$ and 2 for (a) and (b), respectively.

1.5 Parallel MRI

Phased array coil technology has seen significant developments in the last decade [Roemer et al., 1990, Carlson and Minemura, 1993, Sodickson and Manning, 1997]. The introduction of coil systems with multiple receiver coils allowed increasing the signal-to-noise ratio (SNR) of the MR images. Nowadays, phased array coils are most often used to reduce the scan time. This approach is referred to as parallel MRI. Spatially encoding, which is typically performed by applying time-consuming magnetic field gradients, can partially be replaced by spatial sensitivity information rooted in such multichannel coil systems. Parallel imaging thus allows accelerated imaging without losing spatial resolution or image contrast. On the contrary, by skipping a fraction R of the phase-encoding steps, EPI related artifacts, such as geometrical distortions, goes down by the same factor [Bammer et al., 2002]. Obviously, there's no such thing as a free lunch. A reduction in signal-to-noise ratio (SNR) is inherent to parallel MRI (pMRI). Therefore, the acceleration factor R is in clinical experiments most often limited to 2 or 4. In this section, a brief technical overview of SENSitivity Encoding (SENSE) [Pruessmann et al., 1999] and generalized autocalibrating partially parallel acquisitions (GRAPPA) [Griswold et al., 2002], two routinely used pMRI reconstruction techniques, is given. A more general overview is given by Blaimer et al. [2004].

1.5.1 SENSE

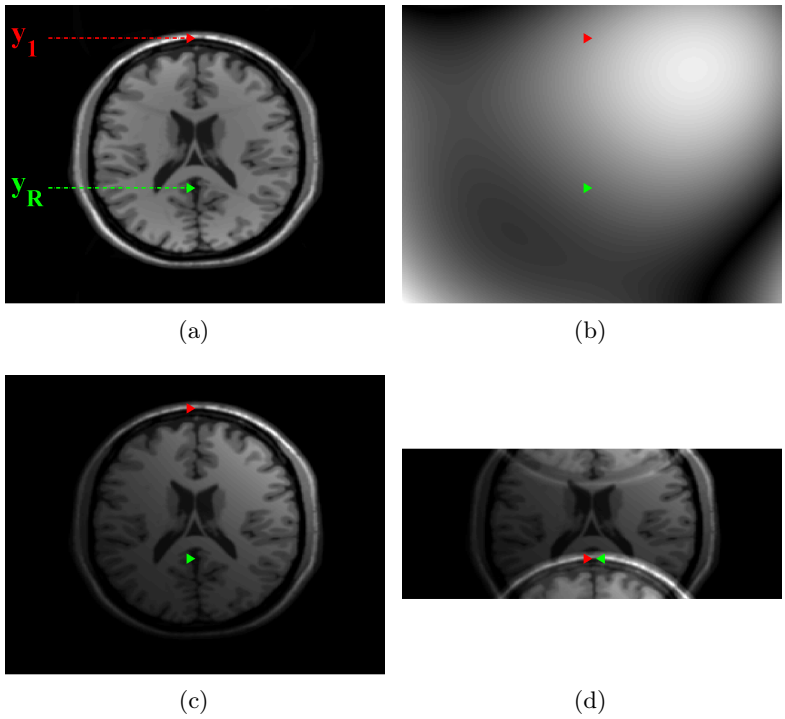
Let's consider a MR receiver system with an array of L, simultaneously operated receiver coils. Furthermore, the coil elements show inhomogeneous, mutually distinct spatial sensitivity. Those sensitivity maps are assumed to be known, or at least, estimated during a calibration process, which often involves the acquisition of additional calibration scans. In pMRI, the number of phase-encoding steps is reduced by the factor R. Hence, the k-space is undersampled. The inverse Fourier transformation of undersampled k-space data yields images with reduced field-of view (FOV), causing aliasing artifacts (see Fig. 1.4(c,d)). More specifically, the FOV is reduced by the factor R. Exploiting the knowledge of the sensitivity maps allows to create an full-FOV image from the L folded images. To achieve this, one must undo the signal superposition underlying the fold-over effect. Indeed, the signal measured at location (x, y) in the k^{th} coil image $\mathcal{C}_k^s$, which has reduced FOV, is the weighted sum of an R-tuple of signals in the full FOV image $\mathcal{C}$:

$$\mathcal{C}_k^s(x, y) = \sum_{i=1}^{R} \mathcal{S}_k(x, y_i)\mathcal{C}(x, y_i), \tag{1.23}$$

with $\{(x, y_i) : i = 1, ..., R\}$ the spatial locations the of involved full FOV pixels and $S_i(x, y_i)$ is the local sensitivity of the k^{th} coil element (see Fig. 1.7 and Fig. 1.8). Eq. (1.23) can be written as follows:

$$\boldsymbol{c}^s = \boldsymbol{S}\boldsymbol{c}, \tag{1.24}$$

with $\boldsymbol{c}^s = [c_1^s, \cdots, c_L^s]^T = [\mathcal{C}_1^s(x, y), \cdots, \mathcal{C}_L^s(x, y)]^T$ being an $L \times 1$ column vector representing the coil image values at location (x, y). Furthermore, $\boldsymbol{S}$ is an $L \times R$



(a) (b)

(c) (d)

Figure 1.7: (a) Actual image; (b) Coil sensitivity map; (c) Observed image with R=1, that is, (a) weighted by (b); (d) Observed image with R=2. In (d), the intensities of (c) at 2 spatial locations (red and green) are superimposed.

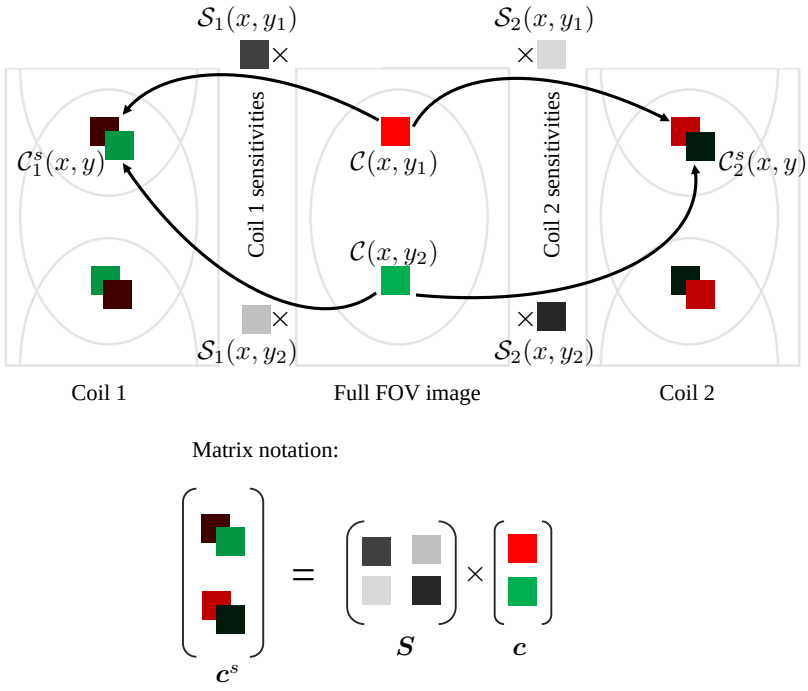


Figure 1.8: Schematic overview of signal superimposition due to undersampling with $R = 2$.

matrix for which $\boldsymbol{S}(i, j) = \mathcal{S}_i(x, y_j)$, and the $R \times 1$ column vector $\boldsymbol{c}$ lists the R pixel values at locations $\{(x, y_i\prime) : i = 1, ..., R\}$ in the full FOV image $\mathcal{C}$. Given the mutually distinct sensitivity maps, and under the condition that $L > R$, one can construct following linear estimator:

$$\hat{\boldsymbol{c}} = \left(\boldsymbol{S}^H \boldsymbol{\Sigma}^{s^{-1}} \boldsymbol{S}\right)^{-1} \boldsymbol{S}^H \boldsymbol{\Sigma}^{s^{-1}} \boldsymbol{c}^s \tag{1.25}$$

with $\boldsymbol{\Sigma}^s$ a covariance matrix, describing the noise characteristics of the different coil elements and H the transposed complex conjugate. The *unfolding* comes with a decrease in SNR:

$$\mathrm{SNR}_{\hat{s}} = \frac{\mathrm{SNR}_s}{g\sqrt{R}}, \tag{1.26}$$

with g a spatially varying geometry factor. The nonuniformity and other noise properties are more thoroughly described in Chapter 3. The reconstruction technique, or closely related alternatives, might also be known as ASSET, SPEEDER or mSENSE. Unlike the others, mSENSE does not require calibration scans to estimate the sensitivity maps. In case of mSENSE, the sensitivity maps are estimated by some additional k-space lines. This is called autocalibration. The additional k-space lines are called autocalibration signal (ACS) lines. The ACS lines can be acquired anywhere in the k-space, however, they are usually acquired near the k-space center because of its high SNR. The reconstruction of ACS lines provide an unaliased, though low resolution, image that can be used for the estimation of

the sensitivity maps. The advantage of autocalibration is the tolerance to patient motion between multiple scans. The downside is a slight decrease in acceleration factor. Autocalibration plays a central role in GRAPPA.

1.5.2 GRAPPA

In GRAPPA, the missing k-space lines are estimated prior to the inverse Fourier transform yielding a full FOV image for each coil element. Filling the missing k-space lines, or in other words, estimating the fully sampled k-space of the k^{th} coil elements $\mathcal{K}_k$, can be seen as a convolution procedure of the acquired undersampled k-spaces $\{\mathcal{K}_i^s : i = 1, ..., L\}$ with a set of GRAPPA reconstruction kernels $\{\boldsymbol{w}_{ki} : i = 1, ..., L\}$ see(Fig. 1.9):

$$\hat{\mathcal{K}}_k = \sum_{i=1}^{L} \mathcal{K}_i^s \otimes \boldsymbol{w}_{ki} \tag{1.27}$$

How to obtain the kernels? $\tilde{\mathcal{K}}_i$ is defined as the fully sampled part of the undersampled k-space $\mathcal{K}_i^s$. Next, $\tilde{\mathcal{K}}_i^s$ is derived from $\tilde{\mathcal{K}}_i$ by nullifying the ACS lines. So, $\tilde{\mathcal{K}}_i^s$ is how $\tilde{\mathcal{K}}_i$ would have looked like without the acquisition of the additional ACS lines. Then, the kernels are the results of following optimizer:

$$\boldsymbol{w}_{ki} = \arg\min_{\boldsymbol{w}_{ki}} \|\tilde{\mathcal{K}}_k - \sum_{i=1}^{L} \tilde{\mathcal{K}}_i^s \otimes \boldsymbol{w}_{ki}\|_2^2. \tag{1.28}$$

Hence, the additionally acquired calibration lines are used to predict the convolution kernels, which are afterwards used to fill up all non-acquired k-space lines. The inverse Fourier transform of $\hat{\mathcal{K}}_k$ results in an unaliased (complex) image $\hat{\mathcal{C}}_k$. After the reconstruction of all L unaliased images, they are commonly combined into a single magnitude image using the sum-of-squares (SoS) formula:

$$\hat{\mathcal{M}} = \sqrt{\sum_{k=1}^{L} \hat{\mathcal{C}}_k^2}. \tag{1.29}$$

Alternatives to the SoS approach are present, but not covered in this work [Gilbert et al., 2007]. Similar to SENSE, a g-factor map indicating the spatially varying SNR reduction can be computed [Breuer et al., 2009]. Again, we refer to Chapter 3 for a more in-depth discussions on noise in MRI.

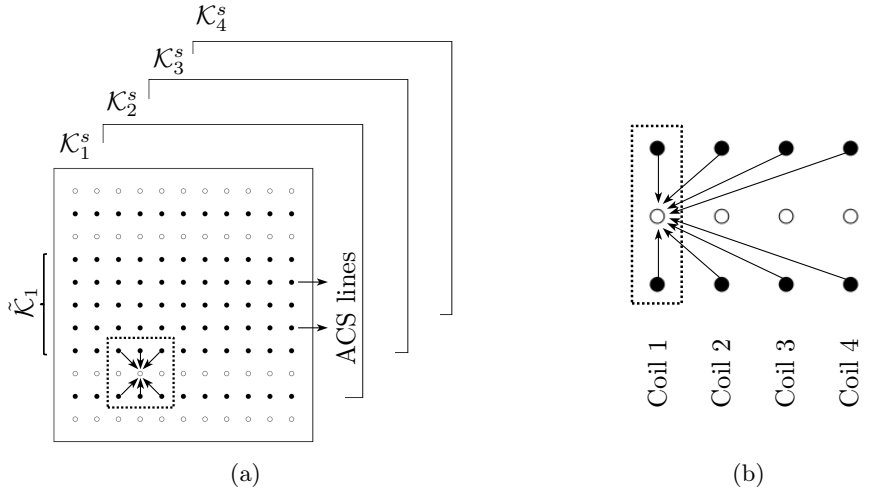


Figure 1.9: Schematic overview of GRAPPA with $L = 4$ and $R = 2$. (a) The acquired k-spaces: the black dots represent acquired k-space data, whereas the white dots were the skipped points. (b) Schematic description of GRAPPA reconstruction: each dot represents a 3-tuple of points in k-space in a single coil of the receiver array. Indeed, the dashed boxes in (a) and (b) represent the same k-space points and arrows. Therefore, each arrow in (b) is basically the sum of three arrows in (a). Multiple weighted k-space points (here 2×3) from all coils are needed to calculate the center point of the 3×3 window coil 1. The weight terms, or the convolution kernels, were estimated a priori by the ACS lines. This procedure needs to be repeated for every point, for every coil, resulting in unfolded coil images, which can be finally combined using – for example – a sum of squares reconstruction.

Bibliography

R. Bammer, M. Auer, S.L. Keeling, M. Augustin, L.A. Stables, R.W. Prokesch, R. Stollberger, M.E. Moseley, and F. Fazekas. Diffusion tensor imaging using single-shot SENSE-EPI. *Magn Reson Med*, 48(1):128–136, 2002.

M. Blaimer, F. Breuer, M. Mueller, R.M. Heidemann, M.A. Griswold, and P.M. Jakob. SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method. *Top Magn Reson Imaging*, 15(4):223–236, 2004.

F. Bloch, W.W. Hansen, and M. Packard. The nuclear induction experiment. *Phys. Rev.*, 70:474–485, 1946.

F.A. Breuer, S.A.R. Kannengiesser, M. Blaimer, N. Seiberlich, P.M. Jakob, and M.A. Griswold. General formulation for quantitative g-factor calculation in GRAPPA reconstructions. *Magn Reson Med*, 62(3):739–746, 2009.

J.W. Carlson and T. Minemura. Imaging time reduction through multiple receiver coil data acquisition and image reconstruction. *Magn Reson Med*, 29(5):681–687, 1993.

H.Y. Carr and E.M. Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.*, 94:630–638, 1954.

R. Damadian. Tumor detection by nuclear magnetic resonance. *Science*, 171(3976): 1151–1153, 1971.

P Dreizen. The nobel prize for MRI: a wonderful discovery and a sad controversy. *The Lancet*, 363(9402):78, 2004.

G. Gilbert, D. Simard, and G. Beaudoin. Impact of an improved combination of signals from array coils in diffusion tensor imaging. *IEEE Trans Med Imaging*, 26(11):1428 –1436, nov. 2007.

M.A. Griswold, P.M. Jakob, R.M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*, 47(6):1202–1210, 2002.

E.L. Hahn. Spin echoes. *Phys. Rev.*, 80:580–594, Nov 1950.

D.I. Hoult, C.N. Chen, and V.J. Sank. Quadrature detection in the laboratory frame. *Magn Reson Med*, 1(3):339–353, 1984.

A. Kumar, D. Welti, and R.R. Ernst. NMR fourier zeugmatography. *J Magn Reson (1969)*, 18(1):69 – 83, 1975.

P.C. Lauterbur. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, 242(5394):190–191, March 1973.

Z.P. Liang and P.C. Lauterbur. *Principles of Magn Reson Imaging: A Signal Processing Perspective*. IEEE Press Series in Biomedical Engineering. SPIE Optical Engineering Press, 2000.

R.J. Macchia, J.E. Termine, and C.D. Buchen. Raymond V. Damadian, M.D.: Magn reson imaging and the controversy of the 2003 nobel prize in physiology or medicine. *The Journal of Urology*, 178(3):783 – 785, 2007.

P. Mansfield. Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):L55, 1977.

J. Mattson and M. Simon. *The pioneers of NMR and Magnc Reson Med: the story of MRI*. Bar-Ilan University Press, 1996.

K.P. Pruessmann, M. Weiger, M.B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magn Reson Med*, 42(5):952–962, 1999.

E.M. Purcell, H.C. Torrey, and R.V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, 69:37–38, Jan 1946.

I.I. Rabi, J.R. Zacharias, S. Millman, and P. Kusch. A new method of measuring nuclear magnetic moment. *Phys. Rev.*, 53:318–318, Feb 1938.

P.B. Roemer, W.A. Edelstein, C.E. Hayes, S.P. Souza, and O.M. Mueller. The NMR phased array. *Magn Reson Med*, 16(2):192–225, 1990.

D.K. Sodickson and W.J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. *Magn Reson Med*, 38 (4):591–603, 1997.

G.J. Stanisz, E.E. Odrobina, J. Pun, M. Escaravage, S.J. Graham, M.J. Bronskill, and R.M. Henkelman. T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magn Reson Med*, 54(3):507–512, 2005.

CHAPTER 2

Diffusion magnetic resonance imaging

Contents

2.1 Introduction

Diffusion magnetic resonance imaging (dMRI) is a popular modality within MRI and currently the only non-invasive method that provides information about the orientation and integrity of tissue – e.g. the brain white matter – microstructure, based on the local water diffusion properties. In this chapter, we cover the basic physics of Brownian motion or self-diffusion. After having described how a spin echo MR sequence can be sensitized to that self-diffusion, we will introduce diffusion tensor imaging and diffusion kurtosis imaging, i.e. two popular diffusion models derived from the cumulant expansion of the diffusion-weighted MR signal.

2.2 Diffusion

2.2.1 Brownian motion

Any type of molecule in a fluid (e.g. water) is in constant motion because of its thermal energy. Given a large collection of molecules in an environment without any obstacles or restrictions, molecules undergo a random walk consisting of independent steps with a change of direction after each collision with another molecule. This motion is named after the botanist Robert Brown, who observed the everlasting jittery motion of grains of pollen suspended in water under a microscope [Brown, 1828]. The Brownian motion was theoretically substantiated by Einstein [1905]. He showed that the random motion of numerous molecules in an open body of water such as a glass of water can statistically be captured by a Gaussian diffusion probability distribution function:

$$p(\boldsymbol{r}) = \frac{1}{\sqrt{(4\pi D\tau)^n}} e^{-\frac{\|\boldsymbol{r}\|^2}{4D\tau}}, \tag{2.1}$$

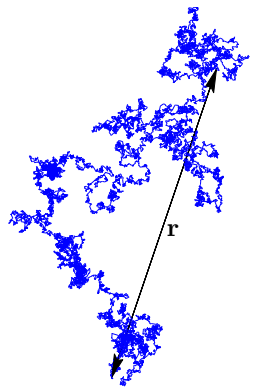


Figure 2.1: A molecule undergoes a random walk consisting of independent steps with a change of direction after each collision with another molecule.

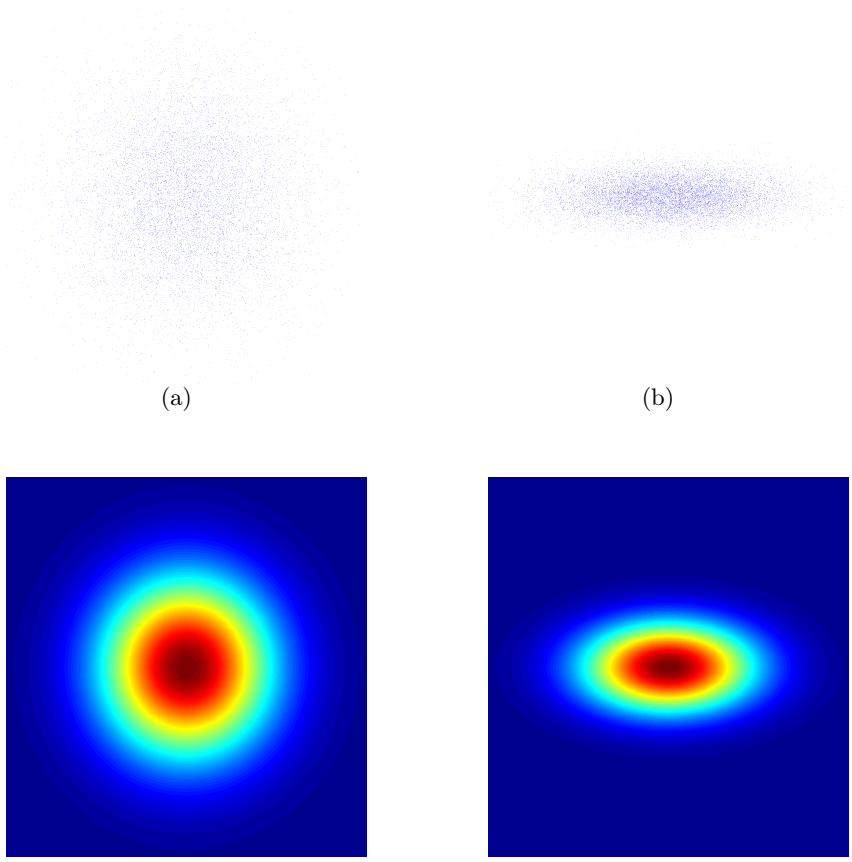
with $\boldsymbol{r}$ an n-dimensional displacement vector, τ the diffusion time, and D the diffusion coefficient, which quantifies the molecules' capacity to diffuse (see Fig. 2.2). In the absence of flow, the Gaussian distribution will be centered around zero. The root-mean-square displacement is given by:

$$r_{\mathrm{RMS}} = \sqrt{\langle \|\boldsymbol{r}\|^2 \rangle} = \sqrt{2nD\tau}, \tag{2.2}$$

with the angle brackets denoting the averaging operator. The metric is often called the characteristic diffusion length or Einstein length. The diffusion length has no dependency on the diffusion direction. Hence, the diffusion process is called isotropic. Obviously a glass of water is a poor model to describe the diffusion in biological tissue. Given typical diffusion times in diffusion-weighted MRI – about 50 to 100 ms – free diffusion can only be expected in the cerebrospinal fluid in the large chambers of the ventricular system. However, biological tissues such as the brain white matter are highly heterogeneous media that consist of various individual compartments (e.g. intracellular, extracellular, neurons, glial cells, and axons) and barriers (e.g. cell membranes and myelin sheaths). Therefore, the random movement of water molecules is hindered and/or restricted by compartmental boundaries and other molecular obstacles. There is no doubt that molecules' mobility is reduced by their interactions with compartments and barriers. However, diffusion is only termed restricted if molecules that are confined in a bounding structure, which they are not likely to leave, collide with this structural boundary during the diffusion time. Typically, the diffusion of water molecules confined within the intra-axonal spaces is expected to be restricted. Indeed, given a diffusion time of 50 ms, a freely diffusing water molecule would displace on average 25 micrometers whereas the diameter of myelinated axons varies between 1 and 20 micrometers. Because of the hindrances and restrictions, the diffusion process in biological tissue can no longer be described by Eq. (2.1). On the one hand, hindrances might reduce the molecules' mobility in particular directions, causing the diffusion process to become anisotropic [Moseley et al., 1991, Chenevert et al., 1990]. On the other hand, restrictions or the presence of multiple diffusion compartments will render the diffusion non-Gaussian (see Fig. 2.3) [Assaf and Cohen, 1998, Beaulieu and Allen, 1994, King et al., 1994, Niendorf et al., 1996, Stanisz et al., 1997, Stanisz and Henkelman, 1998].

2.2.2 Apparent diffusion coefficient and tensor

In biological tissue, molecules' mobility might be reduced by interactions of the diffusing molecules with compartments and barriers. However, the strength of the reduction is time and – possibly – direction dependent. For realistic diffusion times, the apparent root-mean-square displacement, and as such the diffusion coefficient, strongly depend on those interactions of the diffusing molecules with the underlying microstructure, rather than on intrinsic diffusion properties. Therefore, it is convenient to substitute D in Eq. (2.1) and Eq. (2.2) by D_{APP}, the *apparent* diffusion coefficient [Le Bihan et al., 1986]. Next, highly ordered microstructure such as axonal tracts in nervous systems or protein filaments in muscle renders the motion parallel to the structure more likely than perpendicular to it. Hence D_{APP} has directionality and, as such, the diffusion cannot be described adequately by a single



(a) (b)

(c) (d)

Figure 2.2: Given the same initial position, different random walks end up in different end positions (blue dots in (a) and (b) for isotropic and anisotropic diffusion, respectively). The Brownian motion processes can be described by their respective displacement probability density function (PDF) of finding a molecule at a specific position, given the starting point and diffusion time. Free diffusion can be well described by a Gaussian PDF, which is centered around the initial position in the absence of flow (c,d). The width of the Gaussian PDF relates to the diffusion coefficient, which is direction dependent in case of anisotropic diffusion (d).

scalar. A more general PDF is needed to characterize the 3D orientation-dependent water diffusion:

$$p(\boldsymbol{r}) = \frac{1}{\sqrt{(4\pi\tau)^3|\boldsymbol{D}|}} e^{-\frac{\boldsymbol{r}^T \boldsymbol{D}^{-1} \boldsymbol{r}}{4\tau}}, \tag{2.3}$$

with $\boldsymbol{D}$ the (apparent) diffusion tensor:

$$\boldsymbol{D} = \frac{\langle \boldsymbol{r}\boldsymbol{r}^T \rangle}{2\tau} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}, \tag{2.4}$$

which is a full symmetric, positive definite tensor with six independent tensor elements [Basser et al., 1994b].

2.2.3 Apparent kurtosis coefficient and tensor

So far, the apparent diffusion is always assumed to be Gaussian. However, in biological tissue, that might be too strict as an assumption. Indeed, the presence of barriers in biological tissues renders short displacements more probable than long ones with respect to Gaussian diffusion. Moreover, the presence of multiple non-interacting water compartments with differing diffusivities might contribute to the non-Gaussian nature of the diffusion (see Fig. 2.3). In 1D, the deviation to Eq. (2.1) can be quantified by the excess kurtosis K, a dimensionless statistical

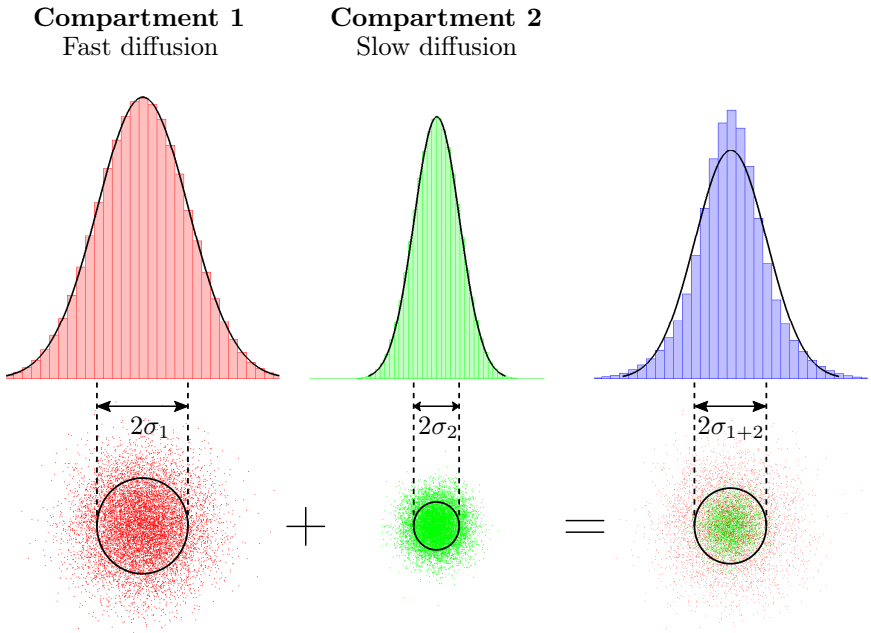


Figure 2.3: A two-compartment model with equal contributions has been used to demonstrate that due to the heterogeneity of the environment the diffusion can no longer assumed to be Gaussian. Given the same initial position (center of black circles), different 2D random walks end up in different end points (red and green points for the fast and slow diffusion compartment, respectively). Histograms of both individual compartments reveal the Gaussian nature of the diffusion. The standard deviation is proportional to the square root of the respective diffusion coefficients. The histogram of the composited compartments shows a clear deviation from a normal distribution (blue bars vs. black line).

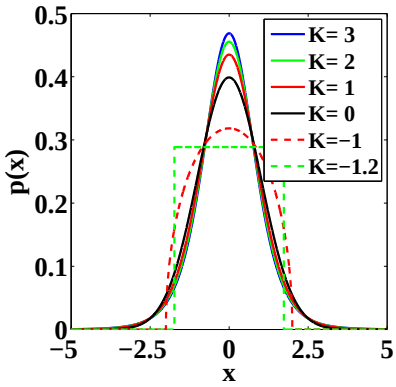


Figure 2.4: Different displacement probability density functions with zero mean, standard deviation one, and varying kurtosis values are shown. Functions with a positive kurtosis are more peaked than the normal distribution (black line)

metric, which quantifies the degree of non-Gaussianity of an arbitrary PDF [Balanda and Macgillivray, 1988]. In case of a zero-centered PDF, the excess kurtosis is the ratio between the fourth and the square of the second moment of the PDF minus three:

$$K = \frac{\langle \boldsymbol{r}^4 \rangle}{\langle \boldsymbol{r}^2 \rangle^2} \tag{2.5}$$

Throughout the thesis, the term excess kurtosis is shortened to kurtosis, although in literature, kurtosis often refers to the first term on the right-hand side. Zero-mean 1D PDFs with equal standard deviation, but varying kurtosis are shown in Fig. 2.4. Basically, the kurtosis measures the peakedness of a distribution. A Gaussian PDF has zero kurtosis, whereas a positive kurtosis indicates that the probabilities of values near the center or extreme values are high in comparison to those of a Gaussian distribution. A PDF with negative kurtosis, on the other hand, has a more rounded peak and wider shoulders, meaning that, compared to a Gaussian distribution, the probability to observe values near the center or for extreme values is lower. The kurtosis has a theoretical lower bound of minus two. Analogue to D_{app}, it is common to refer to K as K_{APP}, the apparent kurtosis [Jensen et al., 2005]. For n-dimensional directional non-Gaussian diffusion, an $n \times n$ covariance matrix – and as such the diffusion tensor – provides an incomplete description of the diffusion process. The (apparent) kurtosis tensor $\boldsymbol{K}$ being the $n-$dimensional generalization of K_{APP} needs to be introduced to quantify the deviation to Gaussianity on a direction dependent basis $\boldsymbol{K}$ is an $n \times n \times n \times n$ tensor, which is fully symmetric with respect to an interchange of indices. For 3D, the tensor has thus 81 tensor elements, of which only 15 elements are independent. In the context of diffusion

MRI, Jensen et al. [2005] introduced the diffusion kurtosis tensor as follows:

$$K_{ijkl} = 9\frac{\langle r_i r_j r_k r_l\rangle - \langle r_i r_j\rangle\langle r_k r_l\rangle - \langle r_i r_k\rangle\langle r_j r_l\rangle - \langle r_i r_l\rangle\langle r_j r_k\rangle}{\langle \boldsymbol{r}^T\boldsymbol{r}\rangle^2}, \tag{2.6}$$

with i, j, k, and l ranging from one to three and $\boldsymbol{r} = [r_1, r_2, r_3]^T = [r_x, r_y, r_z]^T$. The absence of odd-order terms in the numerator suggests the assumption of antipodal diffusion. At this point, it is sufficient to grasp that hindrances and restrictions inherent to the microstructure hampers the statistical description of the displacements of a collection of diffusing molecules. Fortunately, the knowledge of the diffusion tensor and the diffusion kurtosis tensor might reveal *some* information about the underlying microstructure. Before elaborating on both tensors, we will first discuss how they can be measured non-invasively using MRI.

2.3 Diffusion-weighted MRI

2.3.1 Stjeskal-Tanner sequence

Already before the introduction of MR imaging, Stejskal and Tanner [1965] were able to demonstrate that placing a pair of dephasing and rephasing gradients of the same polarity on either side of the 180°-pulse of a classic spin echo sequence causes the echo MR signal to be sensitized by molecular diffusion. Let's consider two rectangular diffusion gradient pulses $\boldsymbol{g}(t)$ with duration time δ and amplitude G:

$$\boldsymbol{g}(t) = G\hat{\boldsymbol{g}}(t), \tag{2.7}$$

with $\hat{\boldsymbol{g}}(t)$ the unit vector denoting the gradient direction. Gradient coils in clinical scanners have maximum gradients magnitudes $G_{\max}$ of $40 - 80\,\mathrm{mT/m}$. The time between the ramp ups of the gradients is Δ. A schematical overview of the so-called *Pulsed Gradient Spin-echo (PGSE) sequence* is given in Fig. 2.5. Owing to the linear relationship between precession rate and the applied magnetic field, the first gradient pulse induces a position-dependent phase shift:

$$\begin{aligned}
\phi_1 &= \gamma \int_0^\delta \boldsymbol{g}(t)\cdot\boldsymbol{x}(t)dt \\
&= \gamma \int_0^\delta G\hat{\boldsymbol{g}}\cdot\boldsymbol{x}(t)dt \\
&= \gamma G \int_0^\delta x_{\boldsymbol{g}}(t)dt \\
&= \gamma\delta G x'_{\boldsymbol{g}}
\end{aligned} \tag{2.8}$$

with $\boldsymbol{x}(t)$ the spin position, and $x_{\boldsymbol{g}}(t)$ the projection of the spin position on the gradient direction. The spin position is assumed to be constant during the pulse duration. This assumption obviously only holds for infinitely short gradient pulses,

i.e. the *short pulse gradient* (SPG) condition. The second gradient induces a similar position-dependent phase shift:

$$\phi_2 = \gamma \int_{\Delta}^{\Delta+\delta} \boldsymbol{g}(t) \cdot \boldsymbol{x}(t) dt = \gamma G \int_{\Delta}^{\Delta+\delta} x_{\boldsymbol{g}}(t) dt = \gamma \delta G x''_{\boldsymbol{g}}, \tag{2.9}$$

The 180°-pulse applied between the two gradient pulses inverts the sign of ϕ_1. Hence, the resulting net dephasing is given by:

$$\phi = \phi_2 - \phi_1 = \gamma \delta G (x''_{\boldsymbol{g}} - x'_{\boldsymbol{g}}) = \gamma \delta G r_{\boldsymbol{g}}, \tag{2.10}$$

with $r_{\boldsymbol{g}} = \boldsymbol{r} \cdot \hat{\boldsymbol{g}}$ the projection of the displacement vector on the gradient direction. The net phase shift is thus zero for static spins. For randomly moving molecules, however, the initial induced phase shift will not fully be cancelled by the second (reversed) phase shift. The incomplete cancellation results in phase incoherence among the spin system, which on its turn, causes a signal drop compared to the observed MR signal in absence of any gradients, i.e. the nondiffusion-weighted signal $S(0)$. The attenuated diffusion-weighted MR signal is given by:

$$\begin{aligned} S(q) &= S(0) \left\langle e^{-iqr_{\boldsymbol{g}}} \right\rangle \\ &= S(0) \int_{-\infty}^{\infty} e^{-iqr_{\boldsymbol{g}}} \bar{p}(r_{\boldsymbol{g}}, \Delta) dr_{\boldsymbol{g}} \leq S(0), \end{aligned} \tag{2.11}$$

with $q = \gamma \delta G$ and $\bar{p}(r, \Delta)$ the ensemble average propagator given by:

$$\bar{p}(r_{\boldsymbol{g}}, \Delta) = \int_{-\infty}^{\infty} \rho(x'_{\boldsymbol{g}}) p(x'_{\boldsymbol{g}}, x'_{\boldsymbol{g}} + r_{\boldsymbol{g}}, \Delta) dx'_{\boldsymbol{g}}, \tag{2.12}$$

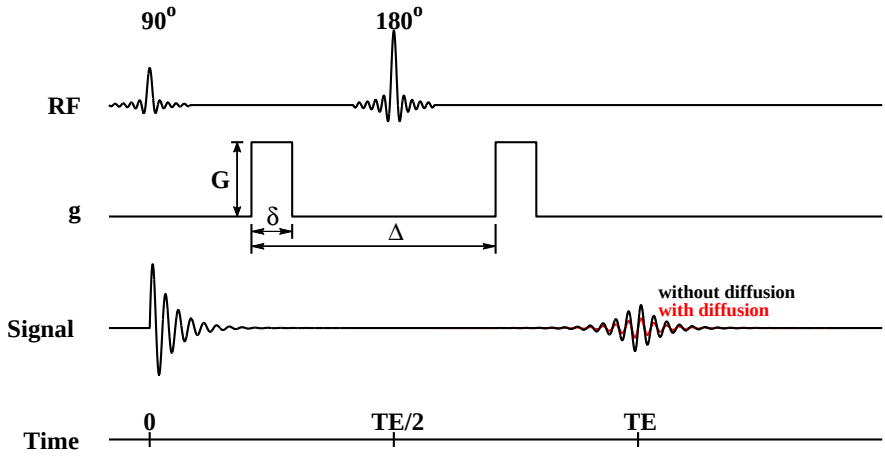


Figure 2.5: Pulsed gradient spin echo sequence

with $p(x'_{\boldsymbol{g}}, x'_{\boldsymbol{g}} + r_{\boldsymbol{g}}, \Delta)$ probability of finding a molecule at specific position $x'_{\boldsymbol{g}} + r_{\boldsymbol{g}}$, given the starting point $x'_{\boldsymbol{g}}$ and diffusion time Δ. Furthermore, $\rho(x'_{\boldsymbol{g}})$ quantifies the probability of finding a spin at location $x'_{\boldsymbol{g}}$ at the start of the dephasing gradient. The ensemble average propagator fully describes the *average* diffusion process of the imaged sample. In the next section, methods trying to extract the PDF – of some of their statistics – out of a set diffusion-weighted signals are discussed.

2.3.2 Q-space imaging

Callaghan et al. [1988] introduced the q-vector $\boldsymbol{q}$ as:

$$\boldsymbol{q} = \frac{\gamma\delta}{2\pi}\boldsymbol{g} = \frac{q}{2\pi}\hat{\boldsymbol{g}}. \tag{2.13}$$

Assuming that the SPG condition holds, Eq. (2.11) can be rewritten as:

$$\frac{S(\boldsymbol{q})}{S(0)} = \int_{-\infty}^{\infty} e^{-i2\pi\boldsymbol{q}\cdot\boldsymbol{r}}\bar{p}(\boldsymbol{r}, \Delta)d\boldsymbol{r}. \tag{2.14}$$

Hence, Eq. (2.14) shows the Fourier relationship between the ensemble average propagator and the normalized diffusion-weighted samples, which are elements of the so-called q-space. The analogy with the k-space is trivial. By acquiring the diffusion signal for a large number of q-values along many different gradient directions the ensemble average propagator can be computed using the Fourier transformation. The q-space needs to be sampled by increasing the gradient strength G and changing the gradient direction. Unfortunately, the limited gradient strength of modern MR scanner and the wish/need for short scan times restrict whole q-space sampling and, as such, the potential of the method. An alternative, though closely related strategy to obtain diffusion properties from q-space samples is given by the cumulant expansion framework.

2.3.3 The cumulant expansion

Let's recall Eq. (2.11). The Taylor series of the natural logarithm of the function in powers of q is given by:

$$\ln\frac{S(q)}{S(0)} = \sum_{n=0}^{\infty} k_n \frac{(-iq)^n}{n!}, \tag{2.15}$$

with k_n the n^{th} order cumulant of the propagator $\bar{p}(r_{\boldsymbol{g}}, \tau)$ describing the diffusion process projected on the gradient direction [Minati and Weglarz, 2007, Kiselev, 2010]. Note that τ – the effective diffusion time – not equals Δ if the SGP condition is violated. In the absence of flow, the diffusion propagator is even w.r.t $r_{\boldsymbol{g}}$. Hence, all odd-order cumulants are zero. The second cumulant equals the variance of the propagator. The variance on its turn links to the apparent diffusion coefficient by the Einstein equation Eq. (2.2):

$$k_2 = 2D_{\text{APP}}\tau, \tag{2.16}$$

whereas the fourth cumulant relates to the apparent kurtosis coefficient:

$$k_4 = K_{\text{APP}} k_2^2 = 4K_{\text{APP}} D_{\text{APP}}^2 \tau^2. \tag{2.17}$$

The cumulant expansion can thus be written as:

$$\ln \frac{S(b)}{S(0)} = -bD_{\text{APP}} + \frac{1}{6} K_{\text{APP}} \left(bD_{\text{APP}}\right)^2 + O(b^3), \tag{2.18}$$

with $b = q^2\tau$ (see section 2.3.4). The cumulant expansion is a method to directly compute diffusion properties such as diffusion coefficient and kurtosis coefficient along the gradient directions without the need for knowing the propagator itself. The widely used diffusion tensor and diffusion kurtosis model both originate in a truncation of this cumulant expansion.

2.3.4 *b*-value

The *b*-value quantifies the sensitivity to diffusion and determines the strength and duration of the diffusion gradient. For an arbitrary diffusion-weighted sequence, the *b*-value can be calculated as:

$$b = \gamma^2 \int_0^{\text{TE}} \left(\int_0^t \tilde{\boldsymbol{g}}(t')dt' \right)^2 dt, \tag{2.19}$$

with $\tilde{\boldsymbol{g}}(t)$ all applied gradients in function of the time [Le Bihan et al., 1986]. For the Stjeskal-Tanner sequence, the *b*-value becomes:

$$b = \gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3} \right). \tag{2.20}$$

2.4 Diffusion Tensor Imaging

2.4.1 DTI model

For Gaussian diffusion, all cumulants except k_2 are zero. The logarithm of the attenuation of the diffusion-weighted signals thus linearly decays with the *b*-value.

$$\ln S(b) = \ln S(0) - bD_{\text{APP}} \tag{2.21}$$

Alternatively, one might say that there is an exponential relation between the diffusion-weighted signal and the apparent diffusion coefficient:

$$S(b) = S(0) e^{-bD_{\text{APP}}} \tag{2.22}$$

To estimate the diffusion coefficient – and potentially the nondiffusion-weighted signal $S(0)$, which might be seen as a second unknown – it is sufficient to acquire two diffusion-weighted signal along the same gradient direction $\hat{\boldsymbol{g}}$, but with different *b*-values. As previously stated, a single scalar cannot describe an anisotropic

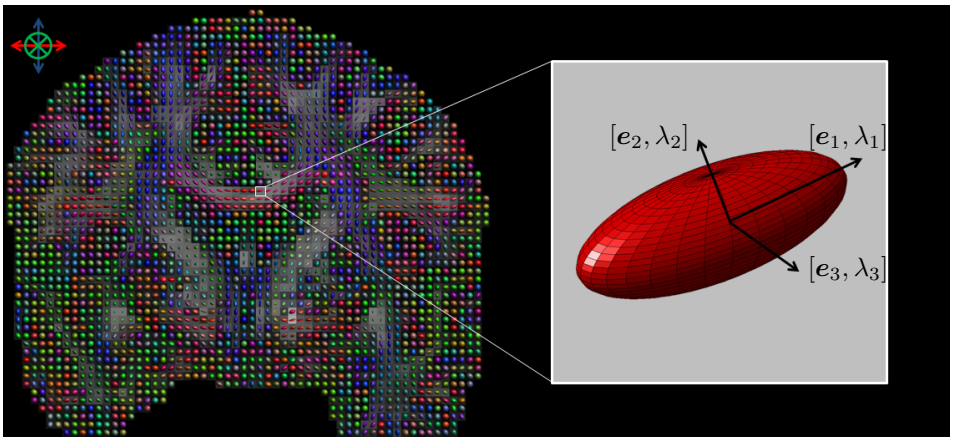


Figure 2.6: The diffusion ellipsoids, color encoded for the diffusion direction – left-right (red), anterioposterior (green), and superior-posterior (blue) – are superimposed on the FA map of a coronal slice. The ellipsoids are uniquely determinded by their eigenvectors and eigenvalues.

Gaussian diffusion process. The apparent diffusion tensor $\boldsymbol{D}$, however, does so. Given that:

$$D_{\text{APP}} = \sum_{i,j=1}^{3} n_i n_j D_{ij} = \hat{\boldsymbol{g}}^T \boldsymbol{D} \hat{\boldsymbol{g}}, \tag{2.23}$$

with D_{ij} the ij^{th} element of $\boldsymbol{D}$, Eq. (2.21) and Eq. (2.22) can be written in terms of $\boldsymbol{D}$:

$$\ln S(b, \hat{\boldsymbol{g}}) = \ln S(0) - b \hat{\boldsymbol{g}}^T \boldsymbol{D} \hat{\boldsymbol{g}}, \tag{2.24}$$

and

$$S(b, \hat{\boldsymbol{g}}) = S(0) e^{-b \hat{\boldsymbol{g}}^T \boldsymbol{D} \hat{\boldsymbol{g}}}, \tag{2.25}$$

respectively. This equation is widely known as the diffusion tensor imaging (DTI) model, originally introduced by Basser and co-workers in the mid-nineties [Basser et al., 1994b].

2.4.2 DTI parameters

According to Eq. (2.4), the diffusion tensor $\boldsymbol{D}$ is basically the 3D covariance matrix of the displacements in a given time. Hence, the diagonal elements correspond to the diffusivities along the three orthogonal axes. The off-diagonal elements, on the other hand, correspond to the correlations between displacements along those axes. The isoprobability surface of the diffusion tensor is an ellipsoid (see Fig. 2.6). The principle axes of the ellipsoid and their corresponding radii, i.e. the diffusion length in a given time along the principal directions, are determined by a decomposition of the diffusion tensor into its real eigenvectors and eigenvalues:

$$\boldsymbol{D} = \boldsymbol{E} \boldsymbol{\Lambda} \boldsymbol{E}^{-1}, \tag{2.26}$$

with $\boldsymbol{E} = [\boldsymbol{e}_1 \quad \boldsymbol{e}_2 \quad \boldsymbol{e}_3]$ being the mutually orthogonal eigenvectors, and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \tag{2.27}$$

the positive eigenvalues for which $\lambda_1 \geq \lambda_2 \geq \lambda_3$ [Hasan et al., 2001]. The eigenvector $\boldsymbol{e}_1$ associated to the largest eigenvalue, is called the principal eigenvector. The direction of the principal eigenvector is assumed to indicate the main diffusion direction, which is for example the direction parallel to the axonal tracts in the brain white matter. The diffusion process is commonly characterized by rotationally invariant scalar measures, calculated from the eigenvalues [Bahn, 1999]:

- **Mean diffusivity (MD)** is the average diffusion coefficient:

$$\mathrm{MD} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}, \tag{2.28}$$

  which can also be computed as the sum of the diagonal tensor elements, i.e. the trace, divided by three.

- **Axial diffusivity (AD)** is the diffusivity along the principal diffusion direction:

$$\mathrm{AD} = \lambda_1. \tag{2.29}$$

- **Radial diffusivity (RD)** is average diffusivity in the equatorial plane, i.e. the plane perpendicular to the principal diffusion direction:

$$\mathrm{RD} = \frac{\lambda_2 + \lambda_3}{2}. \tag{2.30}$$

- **Fractional anisotropy (FA)** is the variance of the eigenvalues, normalized by the magnitude of the tensor [Basser, 1995]:

$$\mathrm{FA} = \sqrt{\frac{3}{2}\frac{(\lambda_1 - \mathrm{MD})^2 + (\lambda_2 - \mathrm{MD})^2 + (\lambda_3 - \mathrm{MD})}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}. \tag{2.31}$$

  The FA can take values between zero (isotropic diffusion) and one (diffusion limited to a single axis).

The orientation of the principal eigenvector is often co-displayed with the FA-images, in separate channels for left-right (red), anterioposterior (green), and superior-posterior (blue) orientations, respectively [Pajevic and Pierpaoli, 1999]. This directionally encoded color (DEC) FA map provide enhanced contrast between different structures.

2.4.3 DTI acquisition

To be able to compute the diffusion tensor with its six independent tensor elements, diffusion-weighting needs to be applied along – at least – six non-collinear and non-coplanar diffusion gradient directions for a single non-zero b-value. In addition, at least, one reference image – typically, though not exclusively, a nondiffusion -weighted image – need to be acquired [Basser et al., 1994a]. The selection of diffusion gradient directions and the strength and number of b-values to maximize precision of the estimated diffusion tensor parameters has been an extensive topic of research. To minimize the amount of noise propagation, it is advised to (a) use the nondiffusion-weighted signal as the reference image, (b) acquire diffusion-weighted images with a b-value around $1000\,\mathrm{s/mm^2}$ [Jones et al., 1999], and (c) oversample the q-space, i.e. acquiring more diffusion-weighted images than the bare minimum [Papadakis et al., 1999]. Indeed, robust and rotationally invariant estimation of the diffusion tensor requires at least 30 gradient directions [Jones, 2004]. Furthermore, the gradient directions are optimally uniformly distributed over a unit sphere. A commonly used algorithm to obtain such an optimal distribution is based on the minimization of electrostatic repulsion [Jones et al., 1999].

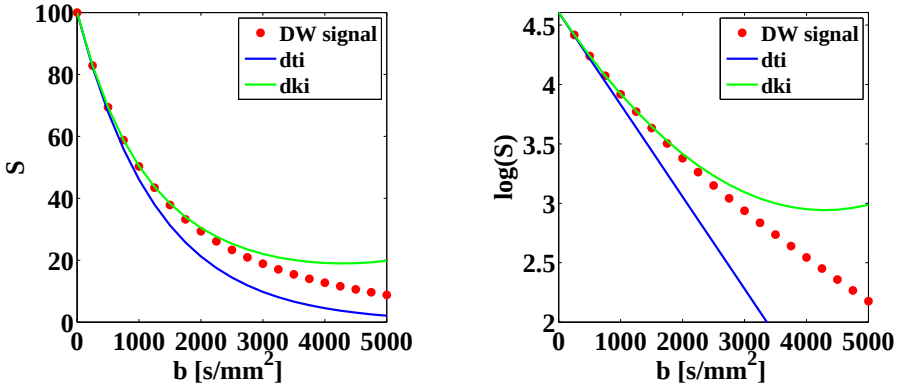
2.4.4 DTI applications

With MD probing the overall water content, and FA indicating the degree of coherence or integrity of underlying structure, various pathological processes at a microscopical level might affect the measured diffusion parameters. Therefore, its widely recognized that DTI can provide an additional insight into the normal and pathological brain. The DTI-derived measures have been applied to study the effects of a continuously growing list of white matter diseases such as multiple sclerosis, Huntington disease, and Alzheimer's disease [Horsfield and Jones, 2002] and psychiatric disorders like schizophrenia [Kubicki et al., 2007]. In addition, diffusion MRI has been used extensively to study brain development [Neil et al., 2002] and aging [Sullivan and Pfefferbaum, 2006]. Another clinical application of DTI is tractography in neurosurgical planning. Tractography is the reconstruction of the pathways of major white matter fiber tracts from the local white matter orientation provided indirectly by the first eigenvector of the diffusion tensor. An accurate localization of those fiber tracts is for example of utmost importance during neurosurgery or to probe brain connectivity. Nowadays different techniques beyond DTI have been presented to bypass the assumption of a single fiber orientation inherent to DTI. An overview of those techniques is outside the scope of this chapter. The interested reader might enjoy the PhD thesis of Ben Jeurissen [2012].

2.5 Diffusion Kurtosis Imaging

2.5.1 DKI model

A natural extension of the diffusion tensor model is given by the diffusion kurtosis model [Jensen et al., 2005, Lu et al., 2006]. The model additionally includes the



(a) (b)

Figure 2.7: Diffusion-weighted signals (left), as well as their log-transformation (right), are shown as a function of the b-value. Owing to the non-Gaussian diffusion, the addition of the b^2-term improves accuracy of the fit. This is mainly noticeable at intermediate b-values. At high b-values, the error term $O(b^3)$ becomes dominant. Therefore, DKI is a low to intermediate b-value technique.

2^{nd} term of Eq. (2.18):

$$\ln \frac{S(b)}{S(0)} = -bD_{\text{APP}} + \frac{1}{6} K_{\text{APP}} \left(bD_{\text{APP}}\right)^2, \tag{2.32}$$

assuming the higher order terms are ignorable. Given Eq. (2.23) and

$$\begin{aligned} K_{\text{APP}} D_{\text{APP}}^2 &= \text{MD}^2 \sum_{i,j,k,l=1}^{3} \hat{g}_i \hat{g}_j \hat{g}_k \hat{g}_l K_{ijkl} \\ &= \left(\sum_{i=1}^{3} \frac{D_{ii}}{3} \right)^2 \sum_{i,j,k,l=1}^{3} \hat{g}_i \hat{g}_j \hat{g}_k \hat{g}_l K_{ijkl}, \end{aligned} \tag{2.33}$$

with K_{ijkl} the $ijkl^{th}$ element of $\boldsymbol{K}$, the apparent kurtosis tensor as defined in Eq. (2.6), one can write Eq. (2.32) as

$$\ln S(b, \hat{\boldsymbol{g}}) = \ln S(0) - b \sum_{i,j=1}^{3} \hat{g}_i \hat{g}_j D_{ij} + \frac{b^2}{6} \left(\sum_{i=1}^{3} \frac{D_{ii}}{3} \right)^2 \sum_{i,j,k,l=1}^{3} \hat{g}_i \hat{g}_j \hat{g}_k \hat{g}_l K_{ijkl}, \tag{2.34}$$

or, equivalently,

$$S\left(b,\hat{\boldsymbol{g}}\right) = S(0)\exp\left[-b\sum_{i,j=1}^{3}\hat{g}_i\hat{g}_j D_{ij} + \frac{b^2}{6}\left(\sum_{i=1}^{3}\frac{D_{ii}}{3}\right)^2\sum_{i,j,k,l=1}^{3}\hat{g}_i\hat{g}_j\hat{g}_k\hat{g}_l K_{ijkl}\right]. \tag{2.35}$$

The diffusion kurtosis model has in total 21 independent tensor elements (6 diffusion tensor elements and 15 diffusion kurtosis tensor elements). Furthermore, most often, $S(0)$ is considered as an additional model parameter.

Typically, a slightly different parameterization of the DKI model is preferred. Indeed, often one substitute $\mathrm{MD}^2\boldsymbol{K}$ by $\tilde{\boldsymbol{K}}$:

$$\ln S(b,\hat{\boldsymbol{g}}) = \ln S(0) - b\sum_{i,j=1}^{3}\hat{g}_i\hat{g}_j D_{ij} + \frac{b^2}{6}\sum_{i,j,k,l=1}^{3}\hat{g}_i\hat{g}_j\hat{g}_k\hat{g}_l\tilde{K}_{ijkl}, \tag{2.36}$$

or, equivalently,

$$S\left(b,\hat{\boldsymbol{g}}\right) = S(0)\exp\left[-b\sum_{i,j=1}^{3}\hat{g}_i\hat{g}_j D_{ij} + \frac{b^2}{6}\sum_{i,j,k,l=1}^{3}\hat{g}_i\hat{g}_j\hat{g}_k\hat{g}_l\tilde{K}_{ijkl}\right], \tag{2.37}$$

with $\tilde{K}_{ijkl}$ the $ijkl^{th}$ element of $\tilde{\boldsymbol{K}}$, the scaled apparent kurtosis tensor. Given this parameterization, the DKI model is like the DTI model log-linear.

2.5.2 DKI parameters

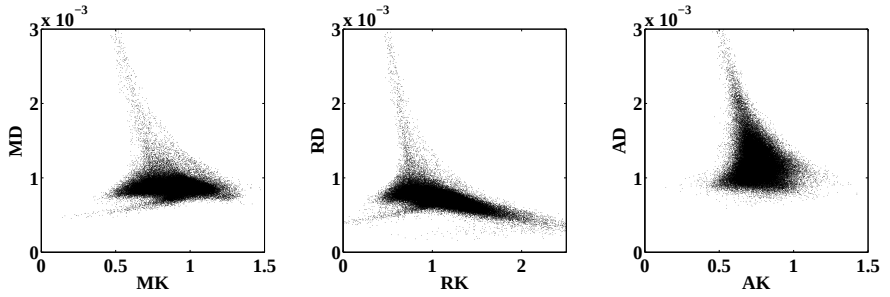
Since the diffusion tensor is a subset of the diffusion kurtosis model parameters, all DTI parameters discussed in section 2.4.2 can again be computed. It is noteworthy that, on the one hand, DKI provides a more objective and accurate quantification of these scalar metrics in that the b-value dependence of the estimated diffusivity is eliminated or at least strongly reduced [Veraart et al., 2011a]. On the other hand, DKI provides additional rotationally invariant metrics of diffusional non-Gaussianity, complementary to the diffusion metric obtained with DTI. The most commonly used kurtosis metrics are:

- **Mean kurtosis (MK)** is the average K_{APP} over a sphere:

$$\mathrm{MK} = \frac{1}{4\pi}\iint_{\Omega} K_{\mathrm{APP}}(\boldsymbol{n})d\Omega, \tag{2.38}$$

  with

$$K_{\mathrm{APP}}(\boldsymbol{n}) = \frac{\left(\sum_{i=1}^{3}\frac{D_{ii}}{3}\right)^2}{\left(\sum_{i,j=1}^{3}n_i n_j D_{ij}\right)^2}\sum_{i,j,k,l=1}^{3}n_i n_j n_k n_l K_{ijkl}, \tag{2.39}$$



(a) (b) (c)

Figure 2.8: Scatter plots show the correlation between (left) mean kurtosis and mean diffusivity, (middle) radial kurtosis and radial diffusivity, and (right) axial kurtosis and axial diffusivity. The corresponding metrics are weakly correlated. The Spearman's rank correlation coefficients are, respectively, -0.07, -0.65, and -0.13.

integrated over the unit sphere Ω with $\boldsymbol{n} = [n_1, n_2, n_3] \in \Omega$ [Lu et al., 2006].

- **Axial kurtosis (AK)** is the apparent kurtosis coefficient, measured along the principal direction $\boldsymbol{e}_1$, determined by the first eigenvector of the diffusion tensor [Hui et al., 2008]:

$$\mathrm{AK} = K_{\mathrm{APP}}(\boldsymbol{e}_1). \tag{2.40}$$

- **Radial kurtosis (RK)** is the average K_{APP}, measured in the equatorial plane, i.e. the plane spanned by $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ [Poot et al., 2010]:

$$\mathrm{RK} = \int_0^{2\pi} K_{\mathrm{APP}}(\boldsymbol{e}_2 \cos(\phi) + \boldsymbol{e}_3 \sin(\phi)) d\phi, \tag{2.41}$$

- **Kurtosis anisotropy (KA)** is in analogy with FA defined as the standard deviation of K_{APP} [Poot et al., 2010]:

$$\mathrm{KA} = \sqrt{\frac{1}{4\pi} \iint_{\Omega} \left(K_{\mathrm{APP}}(\boldsymbol{n}) - \mathrm{MK}\right)^2 d\Omega}, \tag{2.42}$$

In practice, the integrals can only be approximated by a summation over a densely sampled sphere. Alternatively, direct formulae to compute MK, RK, and AK are given in [Tabesh et al., 2011, Jensen and Helpern, 2010]. The complementariness of the kurtosis and diffusion measures is indicated in the scatter plots of Fig. 2.8, which show the weak correlation between directional diffusion and kurtosis metrics, observed in the white matter of the healthy human brain (cf. [Jensen et al., 2005]). Typical values of the DKI metrics for the healthy human brain are presented in [Lätt et al., 2013]. The parameter maps are shown in Fig. 2.8. The diffusion kurtosis metrics are potentially more sensitive to local (microstructural) tissue properties [e.g. Jensen et al., 2005, and section 2.5.4]. Furthermore, it has been shown that

the diffusion kurtosis metrics are less sensitive to certain confounding effects and thereby serve as a more robust biomarker. One study, for example, showed that the mean kurtosis in gray matter is altered substantially less by CSF contamination than either of the conventional diffusion metrics [Yang et al., 2013].

2.5.3 DKI acquisition

Since the apparent diffusion tensor has 6 independent elements and the kurtosis tensor has 15 elements, there is a total of 21 parameters to be estimated. As an additional degree of freedom is associated with the noise-free nondiffusion-weighted signal, S(0), at least 22 diffusion-weighted images must be acquired for DKI. It can be further shown that there must be at least three distinct b-values, which only differ in the gradient magnitude. Furthermore, at least 15 distinct diffusion (gradient) directions are required [Jensen et al., 2005]. Some additional consideration must be made. The maximal b-value should be chosen carefully and is a trade-off between accuracy and precision. While for DTI, diffusion-weighted images are typically acquired with rather low b-values, about 1000 s/mm^2 , somewhat stronger diffusion sensitizing gradients need to be applied for DKI as the quadratic term in the b-value needs to be apparent (see Fig. 2.7). It is shown that b-values of about 2000 s/mm^2 are sufficient to measure the degree of non-Gaussianity with an acceptable precision [Jensen and Helpern, 2010]. Nevertheless, several studies reported b-values up to 3000 s/mm^2 and even more [e.g. Grinberg et al., 2012, Wang et al., 2011]. The assumption that the diffusion-weighted signal is a monotonically decreasing function in terms of the b-value imposes an analytical upper bound on the maximal b-value [Lazar et al., 2008, Tabesh et al., 2011, Veraart et al., 2011b]:

$$b_{\text{MAX}} \leq 3/(D_{\text{APP}} K_{\text{APP}}) \tag{2.43}$$

Typical diffusion and kurtosis values, observed in the human brain, are $D_{\text{APP}} \approx 1\,\mu\text{m}^2/\text{ms}$ and $K_{\text{APP}} \approx 1$. Those values as such justify the use of b-values up to 3000 s/mm^2 for studies involving the human brain [Jensen and Helpern, 2010]. High b-valued diffusion-weighted images suffer from low signal-to-noise ratio due to the severe signal attenuation. Since SNR has a direct impact on the precision and accuracy of the diffusion quantification, the acquisition of dMRI along more diffusion directions than strictly necessary is advisable. In practice, a minimum of 30 directions for each b-value is fairly common [Jensen and Helpern, 2010]. However, note that gradient directions might vary from b-value to b-value. A wide range of DKI data acquisition protocols in line with these considerations are possible and reported in recent literature. Depending on the set of diffusion parameters one is interested in, a specific acquisition protocol can be computed, i.e. b-values and gradient directions, that is optimal in terms of highest achievable precision on the measurements of interest [Poot et al., 2010].

2.5.4 DKI applications

Despite DKI being a recently developed technique, an exponential growth of publications already suggests DKI to become a new important imaging modality in detecting microstructural changes, e.g. following pathological alterations, in human

living tissue that are not revealed by the Gaussian DTI model. Preliminary, though promising results, showed better differentiation between high-grade and low-grade cerebral gliomas [Raab et al., 2010, Van Cauter et al., 2012]. Furthermore, clinical studies indicate that DKI has the potential to improve the early diagnosis of, or to gain more insight in pathologies such as Parkinson Disease [Wang et al., 2011, Giannelli et al., 2012], attention-deficit hyperactivity disorder [Helpern et al., 2011], temporal lobe epilepsy [Gao et al., 2012], traumatic brain injury [Grossman et al., 2012], Alzheimer's disease [Gong et al., 2013], and cerebral infarction [Jensen et al., 2011, Hori et al., 2012, Hui et al., 2012]. Its potential use, however, is not restricted to the brain. The greater relative contrast of kurtosis metrics for cancerous sextants also suggests the potential clinical advantage of incorporating DKI into liver and prostate MR imaging protocols [Rosenkrantz et al., 2012b,a]. Additionally, one study reported on the sensitivity of the DKI metrics to abnormalities in the lung, i.e. the bronchioles and bronchi, using hyperpolarized ^{3}HE imaging [Trampel et al., 2006]. Microstructural changes associated with human development and aging were studied with DKI [Falangola et al., 2008]. The study showed different mean kurtosis patterns for different age ranges, indicating that DKI is able to detect changes in microstructural complexity for both gray and white matter. Complementary, the practical utility of DKI for the (early) diagnosis of pathological changes has been studied in small animal imaging. It was, for example, shown that DKI enhanced the early detection of ischaemic lesion, associated with a stroke model for rats, compared to DTI [Grinberg et al., 2012]. Moreover, DKI may help stratify heterogeneous diffusion-weighted MRI lesions for enhanced characterization of ischemic tissue injury [Cheung et al., 2012]. Furthermore, studies reported on the increased sensitivity of kurtosis metrics to changes in the white and gray matter, associated to rodent models for Huntington Disease [Blockx et al., 2012b,a], Chronicle mild stress [Delgado y Palacios et al., 2011], Alzheimer's disease [Zhang et al., 2012], traumatic brain injury [Zhuo et al., 2012], and brain maturation [Cheung et al., 2009]. It might be expected that many new potential applications of DKI will be revealed in the near future.

2.6 Challenges and limitations

2.6.1 Interpretation

The potential risk of DKI and DTI is the over-interpretation of observed changes in diffusional measures. Both models just arise from a mathematical expansion of the diffusion-weighted signal as a function of the b-value and, as such, do not involve any biophysical modeling [Fieremans et al., 2011, De Santis et al., 2012, Nilsson et al., 2013]. From a change in kurtosis or diffusivity, one might only conclude that there is something in the tissue microstructure that is changing the way that molecules can diffuse. More specific inferences, however, are not substantiated without the justification of a biophysical model that helps to interpret the biophysical meaning of DTI/DKI metric changes. Recently, the two-tensor model has been studied to elucidate the underpinnings of DKI contrast [Fieremans et al., 2011]. In that model, it is assumed that brain white matter consists of two non-exchanging compartments: an intra-axonal space, consisting of parallel

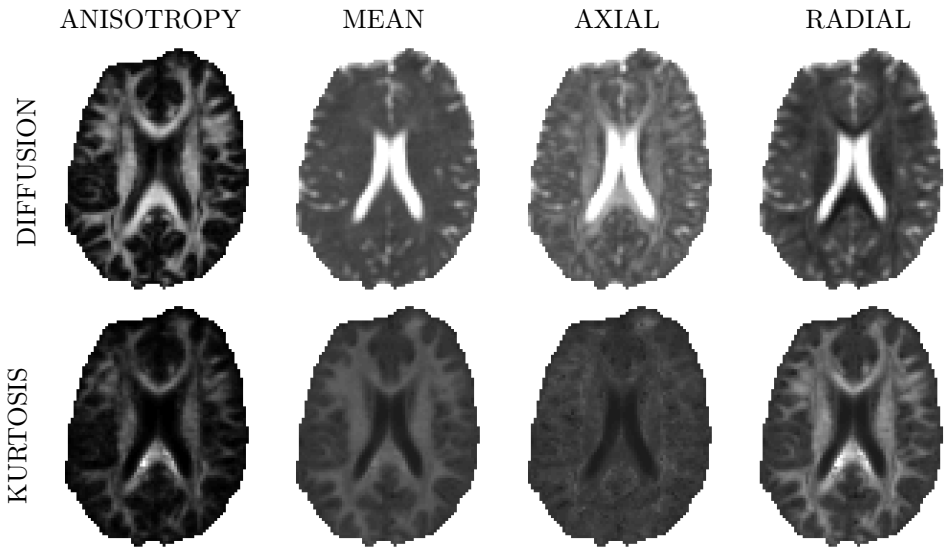


Figure 2.9: The main diffusion and kurtosis parameter maps, obtained from the healthy human brain, are shown. The range of the mean, axial and radial diffusivity is $[0, 3 \times 10^{-3}]\,\mathrm{mm}^2/\mathrm{s}$, while the range of the corresponding kurtosis metrics was [0, 1.5]. The anisotropy maps are bounded by [0, 1].

impermeable cylindrical axons and an extra-axonal space. The diffusion in both compartments is assumed to be anisotropic and Gaussian. The white matter model links the DKI metrics to microstructural properties such as the axonal water fraction and the tortuosity of the extra-axonal space. In another attempt to gain insight in the mathematical DKI model, the DKI information was matched to the information extracted from the biophysical composite hindered and restricted model of diffusion (CHARMED) model [Assaf et al., 2004, Assaf and Basser, 2005, De Santis et al., 2012]. In that model, the white matter is again assumed to be consisting of two compartments: (a) a hindered extra-axonal space, and (b) one or more intra-axonal compartments modeled as impermeable cylinders showing restricted diffusion perpendicular to the fiber. The CHARMED model allows the description of the diffusion weighted-signals in terms of biophysical parameters such as extra- and intra-axonal volume fractions and axonal diffusivities. It was shown that those biophysical parameters correlate with the DKI parameters in areas of higher intra-voxel directional coherence, and as such, the CHARMED model might be used the get more insight into the meaning of the DKI parameters [De Santis et al., 2012]. Those findings, however, only apply within the limits of the validity of both white matter models.

2.6.2 *b*-value dependency of DTI parameters

Diffusion of water molecules is a physical property of the tissue being measured and, thus, its estimated coefficient should not depend on scanner settings or properties, such as the *b*-value. However, due to the non-linear relation between the natural logarithm of the diffusion weighted signal and the *b*-value, the DTI model results in an inaccurate and a *b*-value dependent parameterization of the diffusion process.

Such a b-value dependency might root in inaccurate data statistics (see Chapter 4) or in the complex relation between the diffusion weighted signal and the b-value due to factors such as cerebral perfusion, restricted diffusion, membrane permeability and extra- and intracellular water compartments. In the former case, more advanced parameter estimators might be used (see Chapters 5 and 6). In case of the latter, the b-value dependency of the DTI measures can strongly be reduced by fitting the DKI model to the diffusion-weighted data [Veraart et al., 2011a].

2.6.3 Resolution and signal-to-noise ratio

Diffusion-weighted images suffers from low SNR due to the induced signal attenuation. A low SNR hampers the precise and accurate estimation of the diffusion model parameters (see the next chapter for more details). Increasing the voxel dimensions to increase the SNR is common practice. Typical voxel sizes for human brain diffusion MRI studies are $2 \times 2 \times 2\text{mm}^3$ to $3 \times 3 \times 3\text{mm}^3$, whereas the axonal diameter only goes up to $30\mu m$ [Beaulieu, 2002]. Partial volume effects, or intra-voxel heterogeneity, are inherent to diffusion MRI. This limits the validity of the assumptions made in diffusion models such as DTI [Jeurissen et al., 2012].

2.6.4 Artifacts

(Diffusion-weighted) MR images are vulnerable for several kinds of artifacts [Le Bihan et al., 2006, Jones and Cercignani, 2010]. A brief overview is given.

2.6.4.1 Motion

Diffusion-weighted MR images are sensitized to random motion of water molecules. Random motion will result in signal loss that can be modeled. Unfortunately, coherent motion cannot be avoided, though, it will strongly affect the MR signal. Even if subjects wouldn't move during the MR scan, there will be still localized coherent movement present due to eye motion, swallowing, or cardiac pulsation. Such coherent motion will introduce random phase shifts in each echo readout. Hence, the acquisition of a single diffusion-weighted image using multiple readouts is challenging because the random phase shift will introduce ghosting, blurring and signal loss [e.g. Skare and Andersson, 2001]. As discussed in section 1.4.2, the ss-EPI sequence fully samples the k-space at once, and as such, the sequence is widely used to avoid such imaging artifacts. Although ss-EPI is less sensitive to motion in comparison to standard sequences, it is considerably more sensitive to off-resonance factors such as B_0 field inhomogeneity, chemical shifts, and eddy current effects from fast switching gradients [Skare and Bammer, 2010].

2.6.4.2 Susceptibility artifact

Resonance frequency offsets cause apparent spatial displacement of signals owing to the low bandwith along the phase-encoding directions during an EPI acquisition. Although such frequency offsets might have multiple causes, local magnetic field inhomogeneities are a primary source. Field inhomogeneities are observed at transitions between regions with different magnetic susceptibility, such as air and

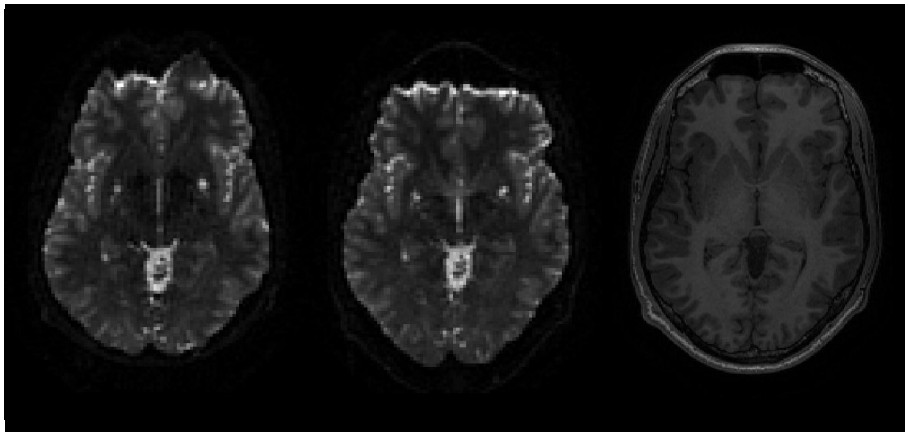


Figure 2.10: Shown are axial EPI axial slices acquired with different phase-encoding directions. The image in right column is the corresponding anatomically undistorted T_1 weighted image. The distortions, and their dependency on the phase-encoding directions, are clearly visible around the frontal sinus.

soft tissue [Jezzard and Balaban, 1995]. Therefore, the susceptibility artifacts are typically observed near the sinuses (see Fig. 2.10) or the auditory canals. The artifact, however, goes beyond geometrical distortions. Due to the non-linearity of the susceptibility-induced distortions, it is possible that the signal intensity from neighboring voxels collapses into a single voxel, resulting in signal drop-outs in the one area and pile-ups in the other [Jones and Cercignani, 2010].

2.6.4.3 Eddy currents

Rapidly switching diffusion gradients will generate eddy currents in nearby conductors. These currents will perturb the spatial encoding locally, and, as such, the reconstructed diffusion-weighted image will be geometrically distorted [Jezzard et al., 1998]. The strength of the distortions – stretch or compression of the image along the phase-encoding direction – increases with the diffusion encoding amplitude [Jones and Cercignani, 2010]. Although the effect of eddy currents can be minimized at the acquisition stage [Reese et al., 2003], residual distortions will still be present. Moreover, those residual distortions vary from one image to another. Correction of those distortions prior to the estimation of the diffusion model parameters is required. A common strategy is to correct subject motion and eddy current distortions simultaneously by a global affine transformation [Netsch and van Muiswinkel, 2004]. However, some criticisms on that approach are ventilated, see e.g. Fig. 2.11. Recently, several techniques dedicated to motion/eddy current correction of high b-valued diffusion-weighted images were presented. Anyhow, the corrections need to be followed by signal modulation according to the volumetric change [Rohde et al., 2004, Jones and Cercignani, 2010]. However, the signal modulation step is widely ignored, partly motivated by a lack of this post-processing step in most software packages.

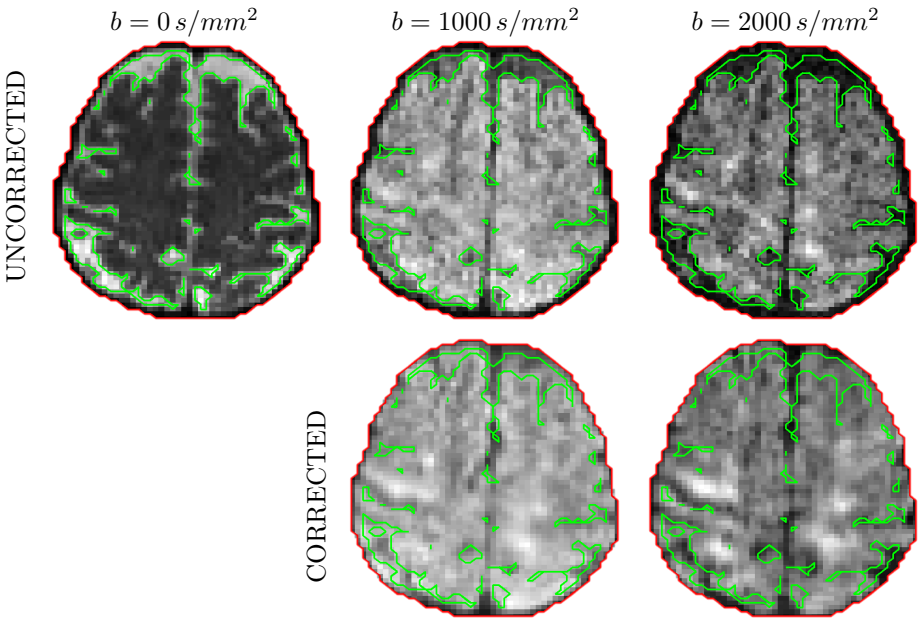


Figure 2.11: Global affine registration based on mutual information seems not to work for high b-values. (top) uncorrected images, (bottom) corrected images using *eddycorrect*. Global affine registrations seems to blow up the brain, ignoring the presence of CSF

2.6.4.4 Chemical shift artifact

The chemical shift artifact is another off-resonance artifact. Protons are magnetically shielded by their surrounding electrons, and as such experience a magnetic field that is different from the external magnetic field. Therefore, the effective precession rates of the proton depend on their chemical environment. Since the chemical environment varies from one molecule to another, protons in fat and water don't have the same resonance frequencies. Given that fat resonates at a slightly lower frequency than water, a mismapping of water and fat signals can be observed [Babcock et al., 1985]. The artifact is clearly recognized by the *fat-band*.

Bibliography

Y. Assaf and P.J. Basser. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage*, 27(1):48–58, 2005. ISSN 1053-8119.

Y. Assaf and Y. Cohen. Non-mono-exponential attenuation of water and N-Acetyl aspartate signals due to diffusion in brain tissue. *J Magn Reson*, 131(1):69 – 85, 1998.

Y. Assaf, R.Z. Freidlin, G.K. Rohde, and P.J. Basser. New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magn Reson Med*, 52(5):965–78, 2004. ISSN 0740-3194.

E.E. Babcock, L. Brateman, J.C. Weinreb, S.D. Horner, and R.L. Nunnally. Edge artifacts in MR images: chemical shift effect. *J Comput Assist Tomogr*, 9(2): 252–257, 1985.

M.M. Bahn. Comparison of scalar measures used in magnetic resonance diffusion tensor imaging. *J Magn Reson*, 139(1):1 – 7, 1999.

K.P. Balanda and H.L. Macgillivray. Kurtosis: A Critical Review. *The American Statistician*, 42(2):111–119, 1988.

P.J. Basser. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR Biomed*, 8(7-8):333–334, 1995.

P.J. Basser, J. Mattiello, and D. Le Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B*, 103(3):247–254, 1994a.

P.J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophys J*, 66(1):259–267, 1994b.

C. Beaulieu. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR Biomed*, 15(7-8):435–455, 2002.

C. Beaulieu and P.S. Allen. Determinants of anisotropic water diffusion in nerves. *Magn Reson Med*, 31(4):394–400, 1994.

I. Blockx, G. De Groof, M. Verhoye, J. Van Audekerke, K. Raber, D. Poot, J. Sijbers, A.P. Osmand, S. Von Hörsten, and A. Van der Linden. Microstructural changes observed with DKI in a transgenic Huntington rat model: evidence for abnormal neurodevelopment. *NeuroImage*, 59(2):957–67, 2012a. ISSN 1095-9572.

I. Blockx, M. Verhoye, J. Van Audekerke, I. Bergwerf, J.X. Kane, R. Delgado Y Palacios, J. Veraart, B. Jeurissen, K. Raber, S. von Hörsten, P. Ponsaerts, J. Sijbers, T.B. Leergaard, and A. Van der Linden. Identification and characterization of Huntington related pathology: an in vivo DKI imaging study. *NeuroImage*, 63 (2):653–62, 2012b. ISSN 1095-9572.

R. Brown. A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosoph Mag*, 4:161–1763, 1828.

P.T. Callaghan, C.D. Eccles, and Y. Xia. NMR microscopy of dynamic displacements: k-space and q-space imaging. *Journal of Physics E: Scientific Instruments*, 21(8):820, 1988.

T.L. Chenevert, J.A. Brunberg, and J.G. Pipe. Anisotropic diffusion in human white matter: demonstration with MR techniques in vivo. *Radiology*, 177(2): 401–5, 1990.

J.S. Cheung, E. Wang, E.H. Lo, and P. Z. Sun. Stratification of heterogeneous diffusion MRI ischemic lesion with kurtosis imaging: evaluation of mean diffusion and kurtosis MRI mismatch in an animal model of transient focal ischemia. *Stroke*, 43(8):2252–4, 2012. ISSN 1524-4628.

M.M. Cheung, E.S. Hui, K.C. Chan, J.A. Helpern, L. Qi, and E.X. Wu. Does diffusion kurtosis imaging lead to better neural tissue characterization? A rodent brain maturation study. *Neuroimage*, 45(2):386 – 392, 2009.

S. De Santis, Y. Assaf, and D.K. Jones. Using the biophysical CHARMED model to elucidate the underpinnings of contrast in diffusional kurtosis analysis of diffusion-weighted MRI. *Magn Reson Mater Phy*, 25(4):267–76, 2012. ISSN 0968-5243.

R. Delgado y Palacios, A. Campo, K. Henningsen, M. Verhoye, D. Poot, J. Dijkstra, J. Van Audekerke, H. Benveniste, J. Sijbers, O. Wiborg, and A. Van der Linden. Magn Reson Imaging and spectroscopy reveal differential hippocampal changes in anhedonic and resilient subtypes of the chronic mild stress rat model. *Biological psychiatry*, 70(5):449–57, September 2011. ISSN 1873-2402.

A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der physik*, 322(8):549–560, 1905.

M.F. Falangola, J.H. Jensen, J.S. Babb, C. Hu, F.X. Castellanos, A. Di Martino, S. H. Ferris, and J0A. Helpern. Age-related non-Gaussian diffusion patterns in the prefrontal brain. *J Magn Reson Imaging*, 28(6):1345–1350, 2008.

E. Fieremans, J.H. Jensen, and J.A. Helpern. White matter characterization with diffusional kurtosis imaging. *NeuroImage*, 58(1):177–88, September 2011. ISSN 1095-9572.

Y. Gao, Y. Zhang, C. Wong, P. Wu, Z. Zhang, J. Gao, D. Qiu, and B. Huang. Diffusion abnormalities in temporal lobes of children with temporal lobe epilepsy: a preliminary diffusional kurtosis imaging study and comparison with diffusion tensor imaging. *NMR Biomed*, 25(12):1369–77, December 2012. ISSN 1099-1492.

M. Giannelli, N. Toschi, L. Passamonti, M. Mascalchi, S. Diciotti, and C. Tessa. Diffusion kurtosis and diffusion-tensor MR imaging in Parkinson disease. *Radiology*, 265(2):645–6; author reply 646–7, November 2012. ISSN 1527-1315.

N.-J. Gong, C.-S. Wong, C.-C. Chan, L.-M. Leung, and Y.-C. Chu. Correlations between microstructural alterations and severity of cognitive deficiency in Alzheimer's disease and mild cognitive impairment: a diffusional kurtosis imaging study. *Magn Reson Imaging*, in press, January 2013. ISSN 1873-5894.

F. Grinberg, L. Ciobanu, E. Farrher, and N.J. Shah. Diffusion kurtosis imaging and log-normal distribution function imaging enhance the visualisation of lesions in animal stroke models. *NMR Biomed*, 25(11):1295–304, November 2012. ISSN 1099-1492.

E.J. Grossman, Y. Ge, J.H. Jensen, J.S. Babb, L. Miles, J. Reaume, J.M. Silver, R.I. Grossman, and M. Inglese. Thalamus and cognitive impairment in mild traumatic brain injury: a diffusional kurtosis imaging study. *Journal of neurotrauma*, 29 (13):2318–27, September 2012. ISSN 1557-9042.

K.M. Hasan, P.J. Basser, D.L. Parker, and A.L. Alexander. Analytical computation of the eigenvalues and eigenvectors in DT-MRI. *J Magn Reson*, 152(1):41 – 47, 2001.

J.A. Helpern, V. Adisetiyo, M.F. Falangola, C. Hu, A. Di Martino, K. Williams, F.X. Castellanos, and J.H. Jensen. Preliminary evidence of altered gray and white matter microstructural development in the frontal lobe of adolescents with attention-deficit hyperactivity disorder: a diffusional kurtosis imaging study. *J Magn Reson Imaging*, 33(1):17–23, January 2011. ISSN 1522-2586.

M. Hori, S. Aoki, and I. Fukunaga. A new diffusion metric, diffusion kurtosis imaging, used in the serial examination of a patient with stroke. *Acta Radiologica Short Reports*, 1(12):1–3, 2012.

M.A. Horsfield and D.K. Jones. Applications of diffusion-weighted and diffusion tensor mri to white matter diseases: a review. *NMR Biomed*, 15(7-8):570–577, 2002. ISSN 1099-1492.

E.S. Hui, M.M. Cheung, L. Qi, and E.X. Wu. Towards better MR characterization of neural tissues using directional diffusion kurtosis analysis. *Neuroimage*, 42(1): 122 – 134, 2008.

E.S. Hui, E. Fieremans, J.H. Jensen, A. Tabesh, W. Feng, L. Bonilha, M.V. Spampinato, R. Adams, and J.A. Helpern. Stroke assessment with diffusional kurtosis imaging. *Stroke*, 43(11):2968–73, November 2012. ISSN 1524-4628.

J.H Jensen and J.A. Helpern. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR Biomed*, 23(7):698–710, 2010.

J.H. Jensen, J.A. Helpern, A. Ramani, H. Lu, and K. Kaczynski. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magn reson imaging. *Magn Reson Med*, 53(6):1432–40, June 2005.

J.H. Jensen, M.F. Falangola, C. Hu, A. Tabesh, O. Rapalino, C. Lo, and J.A. Helpern. Preliminary observations of increased diffusional kurtosis in human brain following recent cerebral infarction. *NMR Biomed*, 24(5):452–7, June 2011. ISSN 1099-1492.

B. Jeurissen. *Improved analysis of brain connectivity using high angular resolution diffusion imaging*. PhD thesis, University of Antwerp, 2012.

B. Jeurissen, A. Leemans, J.-D. Tournier, D.K. Jones, and J. Sijbers. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magn reson imaging. *Hum Brain Mapp*, May 2012.

P. Jezzard and R.S. Balaban. Correction for geometric distortion in echo planar images from B0 field variations. *Magn Reson Med*, 34(1):65–73, Jul 1995.

P. Jezzard, A.S. Barnett, and C. Pierpaoli. Characterization of and correction for eddy current artifacts in echo planar diffusion imaging. *Magn Reson Med*, 39(5): 801–812, May 1998.

D.K. Jones. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a monte carlo study. *Magn Reson Med*, 51(4):807–815, Apr 2004.

D.K. Jones and M. Cercignani. Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR Biomed*, 23(7):803–20, August 2010.

D.K. Jones, M.A. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magn reson imaging. *Magn Reson Med*, 42 (3):515–525, 1999.

M.D. King, J. Houseman, S.A. Roussel, N. Van Bruggen, S.R. Williams, and D.G. Gadian. q-space imaging of the brain. *Magn Reson Med*, 32(6):707–713, 1994.

V.G. Kiselev. The cumulant expansion: an overarching framework for understanding diffusion MRI. In Derek K Jones, editor, *Diffusion MRI: Theory, Methods and Applications*, chapter 10, pages 152–68. Oxford University Press, Oxford, 2010.

M. Kubicki, R. McCarley, C.-F. Westin, H.-J. Park, S. Maier, R. Kikinis, F.A. Jolesz, and M.E. Shenton. A review of diffusion tensor imaging studies in schizophrenia. *Journal of Psychiatric Research*, 41(1 - 2):15 – 30, 2007.

J. Lätt, M. Nilsson, R. Wirestam, F. Stahlberg, N. Karlsson, M. Johansson, P. C. Sundgren, and D. van Westen. Regional values of diffusional kurtosis estimates in the healthy brain. *J Magn Reson Imaging*, 37(3):610–8, March 2013. ISSN 1522-2586.

M. Lazar, J.H. Jensen, L. Xuan, and J.A Helpern. Estimation of the orientation distribution function from diffusional kurtosis imaging. *Magnc Reson Med*, 60 (4):774–81, October 2008. ISSN 1522-2594.

D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–7, November 1986.

D. Le Bihan, C. Poupon, A. Amadon, and F. Lethimonnier. Artifacts and pitfalls in diffusion MRI. *J Magn Reson Imaging*, 24:478–488, 2006.

H. Lu, J.H. Jensen, A. Ramani, and J.A. Helpern. Three-dimensional characterization of non-Gsussian water diffusion in humans using diffusion kurtosis imaging. *NMR Biomed*, 19(2):236–247, 2006.

L. Minati and W.P. Weglarz. Physical foundations, models, and methods of diffusion magnetic resonance imaging of the brain: A review. *Concepts in Magnetic Resonance Part A*, 30A(5):278–307, September 2007.

M.E. Moseley, J. Kucharczyk, H.S. Asgari, and D. Norman. Anisotropy in diffusion-weighted MRI. *Magn Reson Med*, 19(2):321–6, June 1991. ISSN 0740-3194.

J. Neil, J. Miller, P. Mukherjee, and P. S. Hï£¡ppi. Diffusion tensor imaging of normal and injured developing human brain - a technical review. *NMR Biomed*, 15(7-8):543–552, 2002.

T. Netsch and A. van Muiswinkel. Quantitative evaluation of image-based distortion correction in diffusion tensor imaging. *IEEE Trans Med Imaging*, 23(7):789–798, Jul 2004.

T. Niendorf, R.M. Dijkhuizen, D.G. Norris, M. van Lookeren Campagne, and K. Nicolay. Biexponential diffusion attenuation in various states of brain tissue: Implications for diffusion-weighted imaging. *Magn Reson Med*, 36(6):847–857, 1996.

M. Nilsson, D. van Westen, F. Stahlberg, P.C. Sundgren, and J. Lätt. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. *Magn Reson Mater Phy*, in press, February 2013. ISSN 0968-5243.

S. Pajevic and C. Pierpaoli. Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: Application to white matter fiber tract mapping in the human brain. *Magn Reson Med*, 42(3):526–540, 1999.

N.G. Papadakis, D. Xing, and C.L.H. Huang. A comparative study of acquisition schemes for diffusion tensor imaging using MRI. *J Magn Reson*, 137(1), 1999. 67-82.

D.H.J. Poot, A.J. Den Dekker, E. Achten, M. Verhoye, and J. Sijbers. Optimal experimental design for diffusion kurtosis imaging. *IEEE Transactions on Medical Imaging*, 29(3):819–829, March 2010. ISSN 1558-254X.

P. Raab, E. Hattingen, K. Franz, F.E. Zanella, and H. Lanfermann. Cerebral gliomas: Diffusional kurtosis imaging analysis of microstructural differences. *Radiology*, 254(3):876, 2010. ISSN 0033-8419.

T.G. Reese, O. Heid, R.M. Weisskoff, and V.J. Wedeen. Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo. *Magnc Reson Med*, 49(1):177–82, January 2003. ISSN 0740-3194.

G.K. Rohde, A.S. Barnett, P.J. Basser, S. Marenco, and C. Pierpaoli. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. *Magn Reson Med*, 51(1):103–114, 2004.

A.B. Rosenkrantz, E.E. Sigmund, G. Johnson, J.S. Babb, T.C. Mussi, J. Melamed, S.S. Taneja, V.S. Lee, and J.H. Jensen. Prostate cancer: feasibility and preliminary experience of a diffusional kurtosis model for detection and assessment of aggressiveness of peripheral zone cancer. *Radiology*, 264(1):126–35, 2012a.

A.B. Rosenkrantz, E.E. Sigmund, A. Winnick, B.E. Niver, B. Spieler, G.R. Morgan, and C.H. Hajdu. Assessment of hepatocellular carcinoma using apparent diffusion coefficient and diffusion kurtosis indices: preliminary experience in fresh liver explants. *Magn Reson Imaging*, 30(10):1534–40, December 2012b. ISSN 1873-5894.

S. Skare and J.L. Andersson. On the effects of gating in diffusion imaging of the brain using single shot EPI. *Magn Reson Imaging*, 19(8):1125–1128, 2001.

S.T. Skare and R. Bammer. EPI-Based Pulse Sequences for Diffusion Tensor MRI. In Derek K Jones, editor, *Diffusion MRI: Theory, Methods and Applications*, chapter 12, pages 182–202. Oxford University Press, Oxford, 2010.

G.J. Stanisz and R.M. Henkelman. Diffusional anisotropy of T2 components in bovine optic nerve. *Magn Reson Med*, 40(3):405–410, 1998.

G.J. Stanisz, G.A. Wright, R.M. Henkelman, and A. Szafer. An analytical model of restricted diffusion in bovine optic nerve. *Magn Reson Med*, 37(1):103–111, 1997.

E.O. Stejskal and J.E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *J Chem Phys*, 42(1):288–292, 1965.

E.V. Sullivan and A. Pfefferbaum. Diffusion tensor imaging and aging. *Neuroscience and Biobehavioral Reviews*, 30(6):749 – 761, 2006.

A. Tabesh, J.H. Jensen, B.A. Ardekani, and J.A. Helpern. Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. *Magn Reson Med*, 65(3):823–36, 2011.

R. Trampel, J.H. Jensen, R.F. Lee, I. Kamenetskiy, G. McGuinness, and G. Johnson. Diffusional kurtosis imaging in the lung using hyperpolarized 3He. *Magn Reson Med*, 56(4):733–7, 2006. ISSN 0740-3194.

S. Van Cauter, J. Veraart, J. Sijbers, R.R. Peeters, U. Himmelreich, F. De Keyzer, S.W. Van Gool, F. Van Calenbergh, S. De Vleeschouwer, W. Van Hecke, and S. Sunaert. Gliomas: Diffusion kurtosis MR imaging in grading. *Radiology*, 263 (2):492–501, 2012.

J. Veraart, D.H.J. Poot, W. Van Hecke, I. Blockx, A. Van der Linden, M. Verhoye, and J. Sijbers. More accurate estimation of diffusion tensor parameters using diffusion Kurtosis imaging. *Magn Reson Med*, 65(1):138–45, 2011a. ISSN 1522-2594.

J. Veraart, W. Van Hecke, and J. Sijbers. Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a Rician noise model. *Magn Reson Med*, 66(3):678–86, 2011b. ISSN 1522-2594.

J.-J. Wang, W.-Y. Lin, C.-S. Lu, Y.-H. Weng, S.-H. NG, C.-H. Wang, H.-L. Liu, R.n-H. Hsieh, Y.-L. Wan, and Y.-Y. Wai. Parkinson disease: diagnostic utility of diffusion kurtosis imaging. *Radiology*, 261(1):210–7, 2011.

A.W. Yang, J.H. Jensen, C.C. Hu, A. Tabesh, M.F. Falangola, and J.A. Helpern. Effect of cerebral spinal fluid suppression for diffusional kurtosis imaging. *J Magn Reson Imaging*, 37(2):365–71, 2013. ISSN 1522-2586.

L. Zhang, R. Chang, L.W. Chu, and H.K.F. Mak. Current neuroimaging techniques in Alzheimer's disease and applications in animal models. *Am J Nucl Med Mol Imaging*, 2(3):386–404, 2012.

J. Zhuo, S. Xu, J.L. Proctor, R. J. Mullins, J.Z. Simon, G. Fiskum, and R.P. Gullapalli. Diffusion kurtosis as an in vivo imaging marker for reactive astrogliosis in traumatic brain injury. *NeuroImage*, 59(1):467–77, 2012. ISSN 1095-9572.

CHAPTER 3

MR data distribution

Contents

3.1 Introduction

Noise is an important source of MR image distortion, especially in dMRI where the sensitization to the diffusion process is characterized by a strong signal decay. In this chapter, we will review the main source of noise in the complex-valued MR signals, which is assumed to be additive and normally distributed. However, the calculation of the signal's magnitude to avoid signal drop-outs in the diffusion-weighted images due to random phase shifts, will alter the data distribution. Although the resulting data distribution is often assumed to be Rician, it will depend on the applied image formation technique. The possible data distributions and their statistics are discussed in this chapter.

3.2 Noise in MRI

Physicist John Johnson experimentally showed that thermal motion of electrons in a resistor, R, results in random fluctuations in the voltage across its terminals [Johnson, 1928]. This voltage fluctuations, known as *Johnson noise* or *thermal noise*, are proportional to the temperature, T $[K]$. The noise variance, or mean-squared noise, is given by:

$$\left\langle V_n^2 \right\rangle = 4k_b T R \Delta f, \tag{3.1}$$

with k_b the Boltzmann constant ($1.38 \times 10^{-23} J/K$) and Δf the receiver frequency bandwidth in Hz [Nyquist, 1928]. Johnson noise was shown to have a uniform spectral density (cfr. white noise), whereas the noise itself is normally distributed. This thermal noise is the principal source of noise in MR imaging. The effective resistance R is the sum of the coil resistance R_c and the resistance induced by the conductive losses in the scanned subject R_s [Hoult and Lauterbur, 1979]. Generally, the latter is the dominant source of noise. At this point, it is sufficient to take up that – in the complex data domain – the noise is additive, stationary, uncorrelated and normally distributed [Henkelman, 1985]. The noise in the corresponding real and imaginary voxels is also assumed to be uncorrelated. Furthermore, the noise has zero mean and variance:

$$\sigma_k^2 = \left\langle V_n^2 \right\rangle. \tag{3.2}$$

Assuming that the noisy k-space data are processed using the linear and orthogonal inverse Fourier reconstruction algorithm, the noise will have the same properties in the (complex) image domain. However, the noise variance will be scaled by the inverse of the number of k-space data points (N_k):

$$\sigma_c^2 = \frac{\sigma_k^2}{N_k}. \tag{3.3}$$

In practice, the story might become more complicated if users (or vendors) choose for example for repeated measurements (NEX) or zero-filling. On the one hand, averaging of repeated measurements to increase the SNR, will reduce the noise variance with a factor NEX. On the other hand, zero-filling, which is often used to increase the apparent spatial resolution or to extend the data length to a power of two, will correlate the noise without changing any of the other properties.

Furthermore, to avoid alias artifacts, low band-pass filtering is applied prior to digitizing the received analogue MR signal. Because of these filters, the actual noise variance will depend on imaging parameters such as FOV in the x and y-directions (FOV_x and FOV_y), number of samples in the x and y-direction of the reconstructed image (N_x and N_y), the sampling interval (Δt), and a factor depending on the filter characteristics [Parker and Gullberg, 1990]:

$$\sigma_c^2 = K \frac{N_x N_y \left\langle V_n^2 \right\rangle}{\mathrm{NEX}\,\mathrm{FOV}_y^2 \mathrm{FOV}_y^2 \Delta t}. \tag{3.4}$$

In that case, the noise variance might become spatially varying along the frequency encoding direction.

Although all information is present in the complex data, it is common practice to compute the magnitude image because they avoid the problem of phase shifts. However, the computation of the magnitude is a nonlinear operation and therefore the noise distribution is no longer Gaussian [Henkelman, 1985, Rice, 1944]. The actual MR magnitude distributions are discussed in the following sections.

3.3 Single-channel acquisition

A single receiver quadrature detector generates a single, complex-valued k-space that can be transformed into a single complex-valued image by the inverse Fourier transform. Both the real and imaginary part of the image is corrupted with zero-mean, normally distributed noise [Gudbjartsson and Patz, 1995]. The noisy image intensity at location $\boldsymbol{x}$ is thus given by:

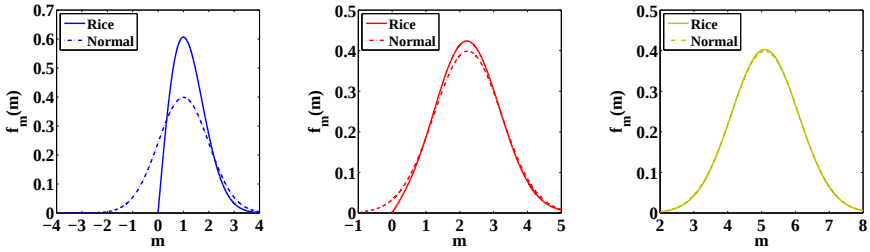
$$\begin{aligned} \mathcal{C}(\boldsymbol{x}) &= \mathcal{C}_0(\boldsymbol{x}) + \epsilon_c(\boldsymbol{x}) \\ &= \mathcal{R}_0(\boldsymbol{x}) + i\mathcal{I}_0(\boldsymbol{x}) + \epsilon_R(\boldsymbol{x}) + i\epsilon_i(\boldsymbol{x}) \\ &= \underbrace{\mathcal{R}_0(\boldsymbol{x}) + \epsilon_r(\boldsymbol{x})}_{\text{real}} + i\underbrace{(\mathcal{I}_0(\boldsymbol{x}) + \epsilon_i(\boldsymbol{x}))}_{\text{imaginary}}, \end{aligned} \tag{3.5}$$

with $\mathcal{R}_0(\boldsymbol{x})$ and $\mathcal{I}_0(\boldsymbol{x})$, respectively, the real and imaginary part of the noise-free complex-valued signal $\mathcal{C}_0(\boldsymbol{x})$. Furthermore, $\epsilon_r(\boldsymbol{x})$ and $\epsilon_i(\boldsymbol{x})$ are the real and imaginary part of the complex-valued noise component $\epsilon_c(\boldsymbol{x})$. Following assumptions regarding the noise terms are made for all $\boldsymbol{x}$:

1. $\epsilon_r(\boldsymbol{x}) \sim \mathcal{N}\left(0, \sigma_c^2\right)$
2. $\epsilon_i(\boldsymbol{x}) \sim \mathcal{N}\left(0, \sigma_c^2\right)$
3. $\mathbb{E}\left[\epsilon_r(\boldsymbol{x})\epsilon_i(\boldsymbol{x})\right] = 0.$

In words, the noise, independently added to both parts of the image, comes from a white noise process with zero mean and standard deviation σ_c. The magnitude image is calculated as the root sum-of-squares (SoS) of the real and imaginary part of the complex image:

$$\mathcal{M}_0(\boldsymbol{x}) = |\mathcal{C}_0(\boldsymbol{x})| = \sqrt{\mathcal{R}_0^2(\boldsymbol{x}) + \mathcal{I}_0^2(\boldsymbol{x})}. \tag{3.6}$$



(a) SNR = 0 (b) SNR = 2 (c) SNR = 5

Figure 3.1: (solid line) PDF of Rice distributed variable with SNR=0, 2 and 5. The SNR varied by changing m_0 whereas σ_c was fixed to 1. (dashed line) Gaussian PDF with the same mean and variance as the Rice distributed variable.

After adding noise, the computed magnitude becomes a random variable:

$$\mathcal{M}(\boldsymbol{x}) = |\mathcal{C}(\boldsymbol{x})| = \sqrt{\left(\mathcal{R}_0(\boldsymbol{x}) + \epsilon_r(\boldsymbol{x})\right)^2 + \left(\mathcal{I}_0(\boldsymbol{x}) + \epsilon_i(\boldsymbol{x})\right)^2}, \tag{3.7}$$

which is best described by its PDF [Bernstein et al., 1989]:

$$f_{\boldsymbol{m}}\left(m|m_0,\sigma_c\right) = \frac{m}{\sigma_c}\exp\left(-\frac{m^2+m_0^2}{2\sigma_c}\right) I_0\left(\frac{m\,m_0}{\sigma_c^2}\right), \tag{3.8}$$

with $m = \mathcal{M}(\boldsymbol{x})$, $m_0 = \mathcal{M}_0(\boldsymbol{x})$, and I_0 the zeroth order modified Bessel function of the first kind [Rice, 1944]. For $m_0 = 0$, the distribution – named after Stephen O. Rice. – reduces to a Rayleigh distribution [Edelstein et al., 1984]:

$$f_m\left(m|0,\sigma_c\right) = \frac{m}{\sigma_c}\exp\left(-\frac{m^2}{2\sigma_c}\right). \tag{3.9}$$

Another asymptotic case of the Rice distribution is the normal distribution:

$$\lim_{m_0\to\infty} f_m\left(m|m_0,\sigma_c\right) = \frac{1}{\sqrt{2\pi}\sigma_c}\exp\left(-\frac{\left(m-m_0\right)^2}{2\sigma_c^2}\right). \tag{3.10}$$

For SNR values – defined as the ratio between m_0 and σ_c – larger than five, the Rice distribution is already well-approximated by a normal distribution with mean $\sqrt{m_0^2+\sigma_c^2}$ and variance σ_c^2 (see Fig. 3.1) [Gudbjartsson and Patz, 1995]:

$$f_m\left(m|m_0,\sigma_c\right) \approx \frac{1}{\sqrt{2\pi}\sigma_c}\exp\left(-\frac{\left(m-\sqrt{m_0^2+\sigma_c^2}\right)^2}{2\sigma_c^2}\right), \tag{3.11}$$

The transition from the Rayleigh distribution to the normal distribution can be observed in Fig. 3.2. For completeness, the moments (and derivatives) of the

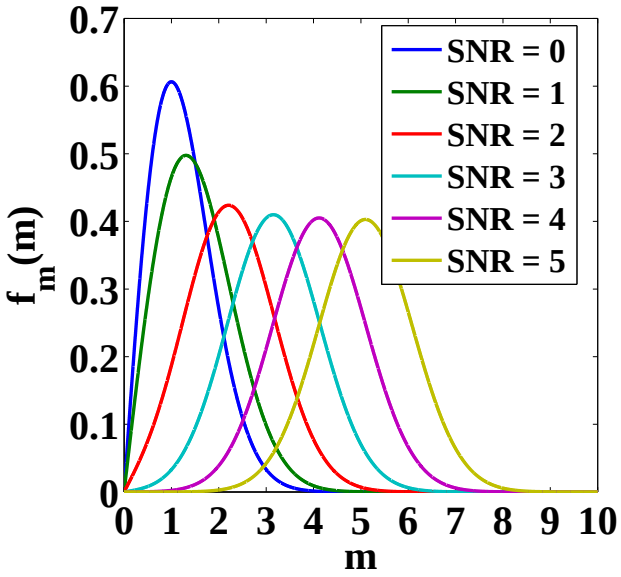


Figure 3.2: PDF of Rice distributed variables with varying m_0 and fixed σ_c. Note that the SNR is defined as the ratio between m_0 and σ_c.

Rice distribution are discussed. The first and second raw moments of the Rice distributed variable are given by

$$\mathbb{E}[m] = \sigma_c \sqrt{\frac{\pi}{2}} L_{1/2}\left(-\frac{m_0^2}{2\sigma_c^2}\right) \tag{3.12}$$

and

$$\mathbb{E}\left[m^2\right] = 2\sigma_c^2 + m_0^2, \tag{3.13}$$

respectively. In Eq. (3.12), $L_{1/2}()$ is the Laguerre polynomial of half order. The first raw moment is better known as the *mean* or *average*. From Eq. (3.12) and Eq. (3.13), the second central moment, also known as the *variance*, can be computed:

$$\begin{aligned} \sigma_m^2 = \mathbb{E}\left[(m - \mathbb{E}[m])^2\right] &= \mathbb{E}\left[m^2\right] - \mathbb{E}[m]^2 \\ &= 2\sigma_c^2 + m_0^2 - \sigma_c^2 \frac{\pi}{2} L_{1/2}^2\left(-\frac{m_0^2}{2\sigma_c^2}\right) \end{aligned} \tag{3.14}$$

The mean and variance of a Rice distributed variable, as a function of the SNR, are shown in Fig. 3.3. The difference between $\mathbb{E}[m]$ and m_0 equals the expectation value of the error term $\epsilon_m = m - m_0$. From a parameter estimation's point of view, it is of utmost importance to grasp that the error term ϵ_m is no longer zero-centered (see next chapter). The skewness and kurtosis are shown as well. Those plots were experimentally obtained. Note that the mean of a Rice distributed variable always exceeds its noise-free value, whereas the variance is always smaller than σ_c^2.

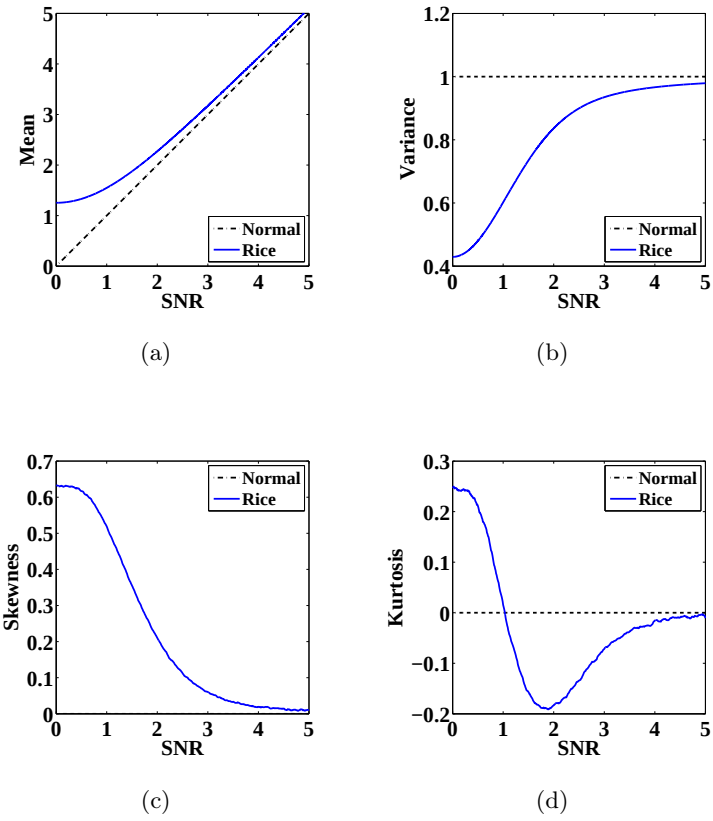


Figure 3.3: Mean (a), variance (b), skewness (c) and kurtosis (d) of a Rice distributed variable as a function of SNR. SNR was changed by changing the underlying signal intensity, while keeping the noise parameter constant, i.e. $\sigma_c = 1$.

3.4 Parallel MRI

Phased array coil technology has seen significant developments in the last decades. The introduction of coil systems with a large number of channels, along with new parallel imaging techniques has resulted in significant improvements in scan times or SNR. The resulting magnitude MR data distribution, however, depends on the reconstruction method that is used to combine the complex signals from all independent channels [Dietrich et al., 2008]. In the next sections, it will be shown that the Rician noise model only holds under well-defined cases.

3.4.1 Unaccelerated MR imaging

If L fully sampled k-spaces are recorded, the sum-of-squares (SoS) method is widely used to combine the individual images into a single composite magnitude image

with improved SNR:

$$\mathcal{M}(\boldsymbol{x}) = \sqrt{\sum_{k=1}^{L} |\mathcal{C}_k(\boldsymbol{x})|^2}, \tag{3.15}$$

with L the number of receiver channels and $\mathcal{C}_k$ the complex image from the k^{th} channel. All $\mathcal{C}_k$ are assumed to be corrupted with zero-mean complex additive Gaussian noise. If, in addition, the variance of noise at each coil is the same, and the signals from the different coils are not correlated, then $\mathcal{M}(\boldsymbol{x})$ follows a noncentral χ distribution, described by following PDF:

$$p(m|m_0) = \frac{m^L}{\sigma_c^2} m_0^{1-L} \exp\left(-\frac{m_0^2 + m^2}{2\sigma_c^2}\right) I_{L-1}\left(\frac{m_0 m}{\sigma_c^2}\right), \tag{3.16}$$

with $m = \mathcal{M}(\boldsymbol{x})$, $m_0 = \mathcal{M}_0(\boldsymbol{x})$ the noise-free composite magnitude signal and I_{L-1} the $(L-1)^{\text{th}}$-order modified Bessel function of the first kind [Constantinides et al., 1997]. The distribution has $2L$ degrees of freedom. Furthermore, note that for $L = 1$, the distribution reduces to the Rice distribution. The PDF is shown for $L = 1, 2, 4, 8, 16$ and 32 in Fig. 3.4. The mean of the noncentral χ distributed

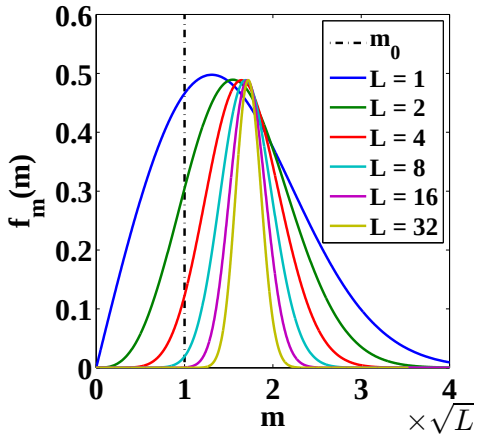


Figure 3.4: PDFs of noncentral χ distributed variables with varying L.

variable is

$$\mathbb{E}[m] = \beta_L \, {}_1F_1\left(-\frac{1}{2}, L, -\frac{m_0^2}{2\sigma^2}\right)\sigma, \tag{3.17}$$

with

$$\beta_L = \sqrt{\frac{\pi}{2}}\frac{(2L-1)!!}{2^{L-1}(L-1)!} = \sqrt{2}\frac{\Gamma(L+0.5)}{\Gamma(L)}, \tag{3.18}$$

and ${}_1F_1(x)$ the confluent hypergeometric function. The variance of m equals

$$\sigma_m^2 = \mathbb{E}\left[m^2\right] - \mathbb{E}[m]^2 = 2L\sigma^2 + m_0^2 - \beta_L^2 \, {}_1F_1^2\left(-\frac{1}{2}, L, -\frac{m_0^2}{2\sigma^2}\right)\sigma^2. \tag{3.19}$$

The mean, variance, skewness and kurtosis for $L = 1, 2, 4, 8, 16$ and 32 are shown in Fig. 3.5.

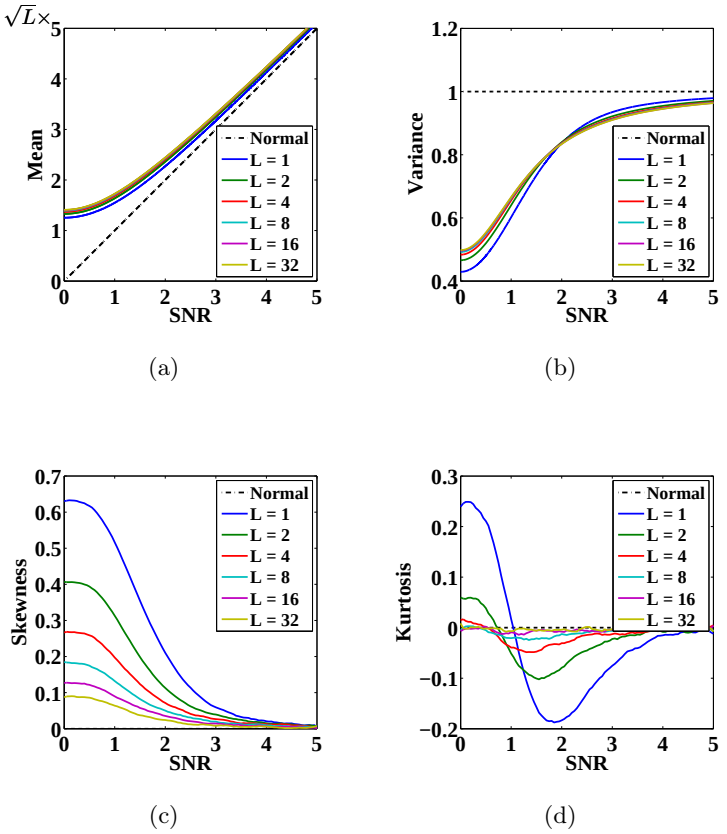


Figure 3.5: Mean (a), variance (b), skewness (c) and kurtosis (d) of a noncentral *chi* distributed variables with varying L as a function of SNR. SNR was changed by varying the underlying signal intensity, while keeping the σ_c constant.

Note that m can only be described by a noncentral χ with $2L$ degrees of freedom if all the receiver antennae in the scanner have the same variance (σ_c^2) of noise and there is no correlation between them. In that case, the covariance matrix $\boldsymbol{\Sigma}$ is a diagonal matrix:

$$\boldsymbol{\Sigma} = \sigma_c^2 \boldsymbol{I}_L, \tag{3.20}$$

with $\boldsymbol{I}_L$ the $L \times L$ identity matrix. However, noise correlations do exist, and they can seriously affect the statistical distribution of data, especially for modern scan systems with a large number of receiving coils. Therefore, in practice, the proposed noncentral χ model is, in general, inaccurate [Aja-Fernández and Tristán-Vega,

2012]. Let us assume a more general covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1L}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \cdots & \sigma_{2L}^2 \\ \vdots & \vdots & & \vdots \\ \sigma_{L1}^2 & \sigma_{L2}^2 & \cdots & \sigma_L^2 \end{bmatrix}. \tag{3.21}$$

The off-diagonal elements stand for the correlations between the respective pair of coils. Although the actual PDF can no longer analytically be expressed, it has been shown that the PDF is still well-approximated by a noncentral χ distribution [Aja-Fernández and Tristán-Vega, 2012]. The number of degrees of freedom will, however, be reduced due to the correlations. Therefore, the composite magnitude data is assumed to be noncentral χ distributed with an *effective parameterization*: effective noise level σ_{eff} and effective number of receiver coils: L_{eff}. Both effective parameters are calculated as:

$$L_{\text{eff}} = \frac{m_0^2 \operatorname{tr}(\boldsymbol{\Sigma}) + (\operatorname{tr}(\boldsymbol{\Sigma}))^2}{\boldsymbol{c}^H \boldsymbol{\Sigma} \boldsymbol{c} + \|\boldsymbol{\Sigma}\|_F^2} \tag{3.22}$$

and

$$\sigma_{\text{eff}}^2 = \frac{\operatorname{tr}(\boldsymbol{\Sigma})}{L_{\text{eff}}}, \tag{3.23}$$

with $\boldsymbol{c} = [\mathcal{C}_1(\boldsymbol{x}), \cdots, \mathcal{C}_L(\boldsymbol{x})]$. As expected, the effective number of coils is lower than the actual one due to the correlations. Furthermore, both effective parameters depend on the underlying signal, and thus, are no longer spatially stationary. However, in case of fully sampled k-space data, or in other words unaccelerated data acquisition, the $\{\sigma_i^2 : i = 1, ..., L\}$ is still assumed to be independent of the spatial location $\boldsymbol{x}$. Therefore, $\operatorname{tr}(\boldsymbol{\Sigma}) = \sum_{i=1}^{L} \sigma_i^2$ is spatially stationary and, as such, $L_{\text{eff}}\sigma_{\text{eff}}^2$ is a constant as well [**?**].

3.4.2 Accelerated MR imaging

The acquisition time is an important issue in a clinical setting. With the evolution of parallel MRI (pMRI), a decrease in acquisition time can be achieved without the need of compromising in spatial resolution or image contrast. pMRI takes full advantage of spatial sensitivity information inherent in an array of multiple receiver coils to partially replace spatial encoding. Indeed, only a fraction of phase-encoding steps needs to be acquired. However, reconstruction methods have to be used to suppress the aliasing and underlying artifacts created by the subsampling of the k-space. The most widely used reconstruction algorithms are SENSE and GRAPPA. Both techniques were already briefly introduced in Chapter 1. In this subsection, a more in-depth description of their implications on the noise properties will be given.

3.4.2.1 SENSE

All aliased complex images $\mathcal{C}_i^s$ with $i = 1, \cdots, L$ are assumed to be corrupted with uniform zero-mean complex additive Gaussian noise. The covariance matrix $\boldsymbol{\Sigma}^s$

describes the levels and correlation of noise in the aliased complex images. Eq. (3.3) showed that the noise variance in the complex images is inversely proportional to the number of k-space samples. A reduction of the number of k-space samples by a factor R, will scale the noise variance by the same factor. Therefore,

$$\boldsymbol{\Sigma}^s = R\boldsymbol{\Sigma}, \tag{3.24}$$

with $\boldsymbol{\Sigma}$ defined as in Eq. (3.21). As explained in section 1.5.1, the SENSE algorithm can be seen as an unfolding technique that reconstructs a single unaliased complex image $\mathcal{C}$ by linearly combining all aliased complex images. Owing to the linearity of the SENSE algorithm, that reconstructed image $\mathcal{C}$ is also disturbed by zero-mean complex additive Gaussian noise. Consequently, each point of the associated magnitude image follows a Rice distribution. The noise variance, however, is no longer spatially uniform [Pruessmann et al., 1999]. Let us recall Eq. (1.25):

$$\hat{\boldsymbol{c}} = \left(\boldsymbol{S}^H \boldsymbol{\Sigma}^{\boldsymbol{s^{-1}}} \boldsymbol{S}\right)^{-1} \boldsymbol{S}^H \boldsymbol{\Sigma}^{\boldsymbol{s^{-1}}} \boldsymbol{c}^s \tag{3.25}$$

Consequently, the $R \times R$ covariance matrix of $\hat{\boldsymbol{c}}$ can be computed as follows:

$$\begin{aligned} \mathrm{cov}(\hat{\boldsymbol{c}}) &= \left[\left(\boldsymbol{S}^H \boldsymbol{\Sigma}^{\boldsymbol{s^{-1}}} \boldsymbol{S}\right)^{-1} \boldsymbol{S}^H \boldsymbol{\Sigma}^{\boldsymbol{s^{-1}}}\right] \boldsymbol{\Sigma}^{s^{-1}} \left[\left(\boldsymbol{S}^H \boldsymbol{\Sigma}^{\boldsymbol{s^{-1}}} \boldsymbol{S}\right)^{-1} \boldsymbol{S}^H \boldsymbol{\Sigma}^{\boldsymbol{s^{-1}}}\right]^H \\ &= \left(\boldsymbol{S}^H \boldsymbol{\Sigma}^{s^{-1}} \boldsymbol{S}\right)^{-1} \end{aligned} \tag{3.26}$$

Since $\boldsymbol{S}^H$ is spatially varying, so will the variance of the reconstructed variables be. Since signals $\hat{\boldsymbol{c}} = \{\hat{\mathcal{C}}(x, y_i) : i = 1, \cdots, R\}$ are basically different linear combinations of the same Gaussian variables, they will be strongly correlated. The standard deviation σ_c^s of the SENSE-reconstructed signal at locations (x, y_i) with $i = 1, \cdots, R$ can be related to the standard deviation σ_c of the signal without subsampling:

$$\sigma_c^s(x, y_i) = \sqrt{R} g(x, y_i) \sigma_c(x, y_i), \tag{3.27}$$

with g a spatially varying geometry factor, which is calculated as:

$$g(x, y_i) = \sqrt{\left[\left(\boldsymbol{S}^H \boldsymbol{\Sigma}^{s^{-1}} \boldsymbol{S}\right)^{-1}\right]_{i,i} \left[\boldsymbol{S}^H \boldsymbol{\Sigma}^{s^{-1}} \boldsymbol{S}\right]_{i,i}} \tag{3.28}$$

3.4.2.2 GRAPPA

In GRAPPA, the missing lines in the subsampled k-spaces are estimated by a linear combination of recorded k-space samples drawn from all coil elements. In Eq. (1.27), completing the k-spaces was written as a convolution of the undersampled k-spaces with a set of GRAPPA reconstruction kernels. Afterwards, L complex images with

full FOV are reconstructed by applying the inverse Fourier transformation:

$$\begin{aligned}\hat{\mathcal{C}}_k &= \mathcal{F}^{-1}\left[\sum_{i=1}^{L} \mathcal{K}_i^s \otimes \boldsymbol{w}_{ki}\right] \\ &= \sum_{i=1}^{L} \mathcal{F}^{-1}\left[\mathcal{K}_i^s \otimes \boldsymbol{w}_{ki}\right] \\ &= \sum_{i=1}^{L} \mathcal{F}^{-1}\left[\mathcal{K}_i^s\right] \times \mathcal{F}^{-1}\left[\boldsymbol{w}_{ki}\right] \\ &= \sum_{i=1}^{L} \mathcal{C}_i^s \times \boldsymbol{W}_{ki},\end{aligned} \tag{3.29}$$

with $\boldsymbol{W}_{ki}$ an inhomogeneous interpolation matrix [Breuer et al., 2009]. The variances and inter-coil correlations for $\{\mathcal{C}_i^s : i = 1, \cdots, L\}$ are again given by $\boldsymbol{\Sigma}^s$, defined in Eq. (3.24). The weighted sum of normally distributed variables is also normally distributed. As such, $\hat{\mathcal{C}}_k$ is disturbed by zero-mean Gaussian noise, whose variance is spatially variable and not constant across the different complex images. Indeed, the covariance matrix of $\hat{\mathcal{C}}_k$ at location (x, y) is given by:

$$\boldsymbol{\Sigma}'(x, y) = \boldsymbol{W}_{ki}(x, y)\boldsymbol{\Sigma}^s \boldsymbol{W}_{ki}^H(x, y), \tag{3.30}$$

Due to the coil-dependency of the noise variance and their correlations, the magnitude data

$$\hat{\mathcal{M}}(x, y) = \sqrt{\sum_{k=1}^{L} |\hat{\mathcal{C}}_k(x, y)|^2}, \tag{3.31}$$

cannot strictly be modeled by a noncentral χ distribution with $2L$ degrees of freedom. Again, an effective parameterization is needed to approximate the actual data distribution with a noncentral χ model [Aja-Fernández et al., 2011]. The effective parameters are now given by:

$$L_{\text{eff}} = \frac{m_0^2 \operatorname{tr}(\boldsymbol{\Sigma}') + (\operatorname{tr}(\boldsymbol{\Sigma}'))^2}{\hat{\boldsymbol{c}}^H \boldsymbol{\Sigma}' \hat{\boldsymbol{c}} + \|\boldsymbol{\Sigma}'\|_F^2} \tag{3.32}$$

and

$$\sigma_{\text{eff}}^2 = \frac{\operatorname{tr}(\boldsymbol{\Sigma}')}{L_{\text{eff}}}, \tag{3.33}$$

with $\hat{\boldsymbol{c}} = \left[\hat{\mathcal{C}}_1(x, y), \cdots, \hat{\mathcal{C}}_L(x, y)\right]$ and $m_0 = \hat{\mathcal{M}}_0(x, y)$. Note that the calculation of the effective parameters requires the knowledge of the noise-free coil images and the GRAPPA kernels. Therefore, it is still very challenging to parametrically describe multichannel MR data if GRAPPA has been used in combination with SoS.

Bibliography

S. Aja-Fernández and A. Tristán-Vega. Influence of noise correlation in multiple-coil statistical models with sum of squares reconstruction. *Magn Reson Med*, 67(2): 580–585, 2012.

S. Aja-Fernández, A. Tristán-Vega, and S.W. Hoge. Statistical noise analysis in GRAPPA using a parametrized noncentral Chi approximation model. *Magn Reson Med*, 65(4):1195–206, 2011. ISSN 1522-2594.

M.A. Bernstein, D.M. Thomasson, and W.H. Perman. Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction. *Med. Phys.*, 16(5):813–817, 1989.

F.A. Breuer, S.A.R. Kannengiesser, M. Blaimer, N. Seiberlich, P.M. Jakob, and M.A. Griswold. General formulation for quantitative g-factor calculation in GRAPPA reconstructions. *Magn Reson Med*, 62(3):739–746, 2009.

C.D. Constantinides, E. Atalar, and E.R. McVeigh. Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magn Reson Med*, 38(5):852–857, 1997.

O. Dietrich, J.G. Raya, S.B. Reeder, M. Ingrisch, M.F. Reiser, and S.O. Schoenberg. Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics. *Magn Reson Imaging*, 26(6):754–762, 2008.

W.A. Edelstein, P.A. Bottomley, and L. M. Pfeifer. A signal-to-noise calibration procedure for NMR imaging systems. *Med. Phys.*, 11(2):180–185, 1984.

H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magn Reson Med*, 34(6):910–914, 1995.

R.M. Henkelman. Measurement of signal intensities in the presence of noise in MR images. *Med Phys*, 12(2):232–233, 1985.

D.I. Hoult and P.C. Lauterbur. The sensitivity of the zeugmatographic experiment involving human samples. *J Magn Reson (1969)*, 34(2):425 – 433, 1979.

J.B. Johnson. Thermal agitation of electricity in conductors. *Phys. Rev.*, 32:97–109, Jul 1928.

H. Nyquist. Thermal agitation of electric charge in conductors. *Phys. Rev.*, 32: 110–113, Jul 1928.

D.L. Parker and G.T. Gullberg. Signal-to-noise efficiency in magn reson imaging. *Med Phys*, 17(2):250–257, 1990.

K.P. Pruessmann, M. Weiger, M.B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magn Reson Med*, 42(5):952–962, 1999.

S.O. Rice. Mathematical analysis of random noise. *Bell Systems Technical Journal*, 23:282–232, 1944.

Part II

Diffusion MRI: parameter estimation

CHAPTER 4

A theoretical introduction to parameter estimation

Contents

4.1 Introduction

The diffusion models, described in Chapter 2, are mathematical relations that describe the diffusion-weighted MR images in terms of diffusion parameters. By using such models, interesting diffusional properties of the object under study can be obtained by estimating the diffusion parameters from a series of acquired diffusion-weighted MR images. However, acquired MR signals are disturbed by noise and are, as such, random variables. The random variable is best described by its probability distribution function (PDF) over the continuous range of its possible outcomes. In Chapter 3, the PDFs describing the intensity of magnitude MR images were discussed. In this chapter, a more general introduction to estimators, i.e. methods to extract information about model parameters from noisy measurements, is given. Popular examples of such estimators are the least squares (LS) estimator and the maximum likelihood (ML) estimator. The general properties of both estimators, their strengths, and limitations will be discussed. Interested readers are referred to van den Bos [2007] for further reading.

4.2 Properties of estimators

4.2.1 Accuracy and precision

We will consider an estimator $\hat{\boldsymbol{\theta}}_N$ of an underlying parameter $\boldsymbol{\theta}_0$ based on N measurements. The accuracy of an estimator is high if the estimated value is *on average* close to the true value. The distance from the average value to the true value, or in other words the systematic estimation error, is called the bias:

$$\text{bias}\left(\hat{\boldsymbol{\theta}}_N\right) = \mathbb{E}\left[\hat{\boldsymbol{\theta}}_N\right] - \boldsymbol{\theta}_0 \tag{4.1}$$

in which $\mathbb{E}[]$ denotes the expectation operator. A *good* estimator has a high accuracy and thus a low, preferably zero, bias. The estimator is called unbiased if the bias is zero for a finite N. Furthermore, the estimator is called asymptotically unbiased if

$$\lim_{N\to\infty} \mathbb{E}\left[\hat{\boldsymbol{\theta}}_N\right] = \boldsymbol{\theta}_0 \tag{4.2}$$

The precision, on the other hand, relates to the *average* spread of the outcomes of the estimator. The precision of an estimator is generally quantified by its variance, i.e. the diagonal elements of the covariance matrix of the estimator:

$$\text{cov}\left(\hat{\boldsymbol{\theta}}_N\right) = \mathbb{E}\left[\left(\hat{\boldsymbol{\theta}}_N - \mathbb{E}\left[\hat{\boldsymbol{\theta}}_N\right]\right)\left(\hat{\boldsymbol{\theta}}_N - \mathbb{E}\left[\hat{\boldsymbol{\theta}}_N\right]\right)^T\right]. \tag{4.3}$$

Note that the estimator's variance is the expectation value of the square of the residual of the estimated values. Obviously, a high precision, or low variance, is also a property of a *good* estimator. An overarching measure to compare the performance, comprising both accuracy and precision, of different estimators is the

mean squared error (MSE):

$$\begin{aligned} \mathrm{MSE}\left(\hat{\boldsymbol{\theta}}_N\right) &= \mathbb{E}\left[\|\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0\|^2\right] \\ &= \sum_i \mathrm{var}(\hat{\boldsymbol{\theta}}_N^{(i)}) + \sum_i \mathrm{bias}(\hat{\boldsymbol{\theta}}_N^{(i)})^2, \end{aligned} \tag{4.4}$$

with $(\cdot)^{(i)}$ the i$^{\text{th}}$ component of the vector.

4.2.2 Properties

Consistency: An estimator is (weakly) consistent if for every $\delta > 0$,

$$\lim_{N\to\infty} \Pr\left[\left\|\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0\right\| > \delta\right] = 0, \tag{4.5}$$

with $Pr\,[]$ being the probabilty. A (weakly) consistent estimator is thus not only asymptotically unbiased. Its variance also converges to zero with $N \to \infty$. The property, however, does not guarantee unbiasedness for finite N.

Efficiency: If within the family of unbiased estimators, $\hat{\boldsymbol{\theta}}_N$ has the highest precision, then $\hat{\boldsymbol{\theta}}_N$ is called efficient. An efficient estimator has thus the smallest MSE, compared to all other unbiased estimators. Note, however, that a biased estimator might have a higher precision.

Normality: If the outcomes of the estimator are normally distributed, the estimator is called normal.

If efficiency or normality is only true for $N \to \infty$, the estimator is called asymptotically efficient or asymptotically normal, respectively.

4.3 Maximum likelihood estimator

Assuming the probability distribution function of variable y is known and given by $p_y(y|\boldsymbol{\theta})$ with $\boldsymbol{\theta}$ as set of parameters, then a general estimation method with optimal (asymptotical) statistical properties, both in terms of accuracy and precision can be developed. The method is known as the maximum likelihood estimator. The maximum likelihood estimator is the maximizer of the likelihood function ($\mathcal{L}$), which has close relation to the probability distribution function. Nevertheless, both functions are fundamentally different. The likelihood function of parameter vector $\boldsymbol{\theta}$ given a measurement y equals the probability of the measurement, given $\boldsymbol{\theta}$:

$$\mathcal{L}\left(\boldsymbol{\theta}|y\right) = p_y(y|\boldsymbol{\theta}) \tag{4.6}$$

As such, the likelihood function is a function of the model parameters, whereas the probability distribution function is a function of the measurements. Loosely speaking, the maximum likelihood estimator tries to determine the model parameters that generate a probability distribution function for which the measurements were

the most probable data. Given a series of measurement $\boldsymbol{y}$ the maximum likelihood estimator maximizes the joint (log)likelihood function:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_N &= \arg\max_{\boldsymbol{\theta}} \prod_{i=1}^{N} \mathcal{L}\left(\boldsymbol{\theta}|y_i\right) \\ &= \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \log \mathcal{L}\left(\boldsymbol{\theta}|y_i\right)\end{aligned} \tag{4.7}$$

Both expressions leads to the same outcome since the logarithmic function is monotonically increasing. However, the maximization of the loglikelihood function is more convenient because nonlinear optimization tools are commonly needed to solve Eq. (4.7). Note that nonlinear optimization tools might suffer from poor convergence due to the local maxima.

If the detailed knowledge of $p_y(y|\boldsymbol{\theta})$ is known, the maximum likelihood estimator hold the attractive property that for N tending to infinity,

$$\hat{\boldsymbol{\theta}}_N \to \mathcal{N}\left(\boldsymbol{\theta}_0, \boldsymbol{J}^{-1}\right), \tag{4.8}$$

with $\boldsymbol{J}$ the Fisher information matrix:

$$J_{ij} = \mathbb{E}\left[-\frac{\partial^2}{\partial \boldsymbol{\theta}_i \boldsymbol{\theta}_j} \log p_y(y|\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] \tag{4.9}$$

The estimator $\hat{\boldsymbol{\theta}}_N$ thus converges in distribution to a normal distribution, centered around $\boldsymbol{\theta}_0$, the vector of noise-free model parameters. Furthermore, the covariance matrix reaches asymptoically the Cramér-Rao lower bound (CRLB), i.e. the minimal achievable variance of all unbiased estimators $\hat{\boldsymbol{\theta}}'_N$:

$$\operatorname{cov}\left(\hat{\boldsymbol{\theta}}'_N\right) \geq \boldsymbol{J}^{-1}. \tag{4.10}$$

Therefore, the maximum likelihood estimator is:

1. asymptotically normal
2. asymptotically unbiased
3. asymptotically efficient
4. consistent

However, one must conclude that there is no guarantee for unbiased and minimum variance estimators for finite N.

In the next section, another popular class of estimators, i.e. the least squares estimators, will be discussed. Furthermore, its relation to the maximum likelihood estimator will be given.

4.4 Least squares estimators

A general form of models describing the relationship between a dependent variable (y) and one or more independent variables ($\boldsymbol{x}$) is

$$y = h\left(\boldsymbol{x}, \boldsymbol{\theta}_0\right) + \epsilon, \tag{4.11}$$

with h an arbitrary function, $\boldsymbol{\theta}_0$ the vector of noise-free model parameters and ϵ an error term with a particular probability distribution function. In general, $\boldsymbol{x}$ is assumed to be measured noise-free, whereas y is noise disturbed. The method of least squares is typically used to obtain an estimate the unknown model parameters, $\hat{\boldsymbol{\theta}}_N$, on the basis of a set of noisy measurements, i.e. $\{(\boldsymbol{x}_i, y_i) : i = 1, ..., N\}$. A least squares estimator results in $\hat{\boldsymbol{\theta}}_N$ that minimizes the sum of squared residual, $\boldsymbol{e}$:

$$\hat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left(y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta})\right)^2 = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \left(e_i\right)^2. \tag{4.12}$$

Note the subtle, though important, difference between *residuals* ($\boldsymbol{e}$) and *errors* ($\boldsymbol{\epsilon}$):

$$\boldsymbol{\epsilon} = \boldsymbol{y} - h(\boldsymbol{x}, \boldsymbol{\theta}_0), \tag{4.13}$$

whereas

$$\boldsymbol{e} = \boldsymbol{y} - h(\boldsymbol{x}, \hat{\boldsymbol{\theta}}_N), \tag{4.14}$$

To study the properties of the least squares estimator, a distinction between linear and nonlinear least squares estimators should be made.

4.4.1 Linear least squares

4.4.1.1 Ordinary linear least squares

Let's assume that the relationship between y and $\boldsymbol{\theta}$ is linear, and as such, can be written as:

$$y = f^{(1)}\left(\boldsymbol{x}\right)\theta_1 + f^{(2)}\left(\boldsymbol{x}\right)\theta_2 + \cdots + f^{(n)}\left(\boldsymbol{x}\right)\theta_n + \epsilon, \tag{4.15}$$

with $f^{(i)}\left(\boldsymbol{x}\right), i = 1 \cdots n$, with n being the number of model parameter, an arbitrary function of $\boldsymbol{x}$. Then, the linear least squares estimator (Eq. (4.12)) can be written in matrix form:

$$\hat{\boldsymbol{\theta}}_N = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}, \tag{4.16}$$

with

$$\boldsymbol{X} = \begin{bmatrix} f^{(1)}\left(\boldsymbol{x}_1\right) & f^{(2)}\left(\boldsymbol{x}_1\right) & \cdots & f^{(n)}\left(\boldsymbol{x}_1\right) \\ f^{(1)}\left(\boldsymbol{x}_2\right) & f^{(2)}\left(\boldsymbol{x}_2\right) & \cdots & f^{(n)}\left(\boldsymbol{x}_2\right) \\ \vdots & \vdots & & \vdots \\ f^{(1)}\left(\boldsymbol{x}_N\right) & f^{(2)}\left(\boldsymbol{x}_N\right) & \cdots & f^{(n)}\left(\boldsymbol{x}_N\right) \end{bmatrix} \quad \text{and} \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \tag{4.17}$$

The design matrix $\boldsymbol{X}$ is an $N \times n$ matrix and the data vector $\boldsymbol{y}$ an $N \times 1$ vector. The solution only exists if X has full column rank. In that case, the inverse of $\left(\boldsymbol{X}^T\boldsymbol{X}\right)$

exists, and the estimator will have a unique solution. Due to the random noise disturbations, the estimator's outcome is also a random variable, which statistical properties must be evaluated to interpret the strengths and limitations of the linear least squares estimator.

Accuracy Substituting $\boldsymbol{y}$ in Eq. (4.16), by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta}_0 + \boldsymbol{\epsilon}, \tag{4.18}$$

results in

$$\hat{\boldsymbol{\theta}}_N = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\left(\boldsymbol{X}\boldsymbol{\theta}_0 + \boldsymbol{\epsilon}\right) = \boldsymbol{\theta}_0 + \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\epsilon}, \tag{4.19}$$

The estimator is unbiased if its expectation value equals the noise-free value, i.e. $\mathbb{E}\left[\hat{\boldsymbol{\theta}}_N\right] = \boldsymbol{\theta}_0$. Assuming that $\boldsymbol{X}$ are not random variables, the equality only holds under the condition of zero-centered error terms, or equivalently, if $\{\mathbb{E}\left[\epsilon_i\right] = 0 : i = 1, ..., N\}$.

Precision The covariance matrix of an unbiased estimator $\hat{\boldsymbol{\theta}}_N$ is given by:

$$\begin{aligned}\operatorname{cov}\left(\hat{\boldsymbol{\theta}}_N\right) &= \mathbb{E}\left[\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0\right)\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0\right)^T\right] \\ &= \mathbb{E}\left[\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right].\end{aligned} \tag{4.20}$$

If $\boldsymbol{\epsilon}$ is a vector of independent, identically and zero-mean distributed error terms, then

$$\mathbb{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right] = \sigma^2\boldsymbol{I_N}, \tag{4.21}$$

with $\boldsymbol{I_N}$ the $N \times N$ identity matrix and σ^2 the variance of $\{\epsilon_i : i = 1, ..., N\}$, then Eq. (4.20) simplifies to:

$$\operatorname{cov}\left(\hat{\boldsymbol{\theta}}_N\right) = \sigma^2\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}. \tag{4.22}$$

Hence, Eq. (4.22) only holds if $\boldsymbol{\epsilon}$ is homoscedastic, i.e. all random variables in the vector have the same variance σ^2. If this assumption does not hold, or equivalently if $\boldsymbol{\epsilon}$ is heteroscedastic, the linear least squares estimator as defined in Eq. (4.16) cannot be the best linear unbiased estimator, i.e. the most precise estimator within the class of unbiased linear estimators. By taking the varying variances into account, the estimator will obtain improved statistical properties.

4.4.1.2 Weighted linear least squares

In linear least squares estimator, the sum of squared residuals is minimized. Hereby, all residuals are weighted equally. However, if the assumption of homoscedasticity does not hold, the introduction of different weighting for each measurement will allow the design of a more precise estimator – the weighted linear least squares estimator. The closed-form solution of the weighted linear least squares estimator is:

$$\hat{\boldsymbol{\theta}}_N = \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}, \tag{4.23}$$

wih $\boldsymbol{W}$ an $N \times N$ weight matrix. Note that if $\boldsymbol{W} = \boldsymbol{I}_N$, Eq. (4.23) reduces to Eq. (4.16). Again, the accuracy and precision of the estimator needs to be evaluated.

Accuracy Analogue to Eq. (4.19), the weighted linear least squares estimator can be rewritten as:

$$\hat{\boldsymbol{\theta}}_N = \boldsymbol{\theta}_0 + \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{\epsilon}. \tag{4.24}$$

Under the assumption that $\boldsymbol{W}$ is deterministic, the estimator is still unbiased if all error terms have zero expectation (cf. linear least squares estimator).

Precision The covariance matrix of $\hat{\boldsymbol{\theta}}_N$, assumed to be unbiased, is now given by:

$$\operatorname{cov}\left(\hat{\boldsymbol{\theta}}_N\right) = \mathbb{E}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\boldsymbol{W}\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\right]. \tag{4.25}$$

If $\boldsymbol{W} = \mathbb{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right]^{-1}$, i.e. the inverse covariance matrix of the error terms, Eq. (4.25) reduces to:

$$\operatorname{cov}\left(\hat{\boldsymbol{\theta}}_N\right) = \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}. \tag{4.26}$$

It has been proven that the weighted linear least squares estimator with this particular choice of weights is the best linear unbiased estimator.

4.4.2 Nonlinear least squares

Often, function h in Eq. (4.11) cannot be expressed as a linear combination of the model parameters (cf. Eq. (4.15)). In that case, model parameters can no longer be determined by linear least squares estimators. There is a need for an alternative, more general method of estimation: nonlinear least squares. The nonlinear least squares estimator is defined as the minimizer of the sum of squared residuals (Eq. (4.12)). The estimator's outcome $\hat{\boldsymbol{\theta}}_N$ meets:

$$\sum_{i=1}^{N}\left[y_i - h(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}}_N)\right]\frac{\partial h(\boldsymbol{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_N\right) = 0. \tag{4.27}$$

The solution must be calculated iteratively. Compared to closed-form linear counterparts, the nonlinear least squares estimator is prone to get stuck in local optima rather than reaching the global optimum. Under some more general assumptions, it can be shown that the nonlinear least squares estimator is asymptotically unbiased if:

1. **Conditional mean:** if $h(\boldsymbol{x}_i, \boldsymbol{\theta}_0) = \mathbb{E}\left[y_i|\boldsymbol{x}_i\right]$ for $i = 1, \cdots, N$. Given Eq. (4.11), it follows from this condition that the error term will again be zero-centered.

2. **Uniqueness:** the estimator has a unique solution, or equivalently, there does not exist a parameter vector $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ such that $h(\boldsymbol{x}_i, \boldsymbol{\theta}) = h(\boldsymbol{x}_i, \boldsymbol{\theta}_0)$ for all $\boldsymbol{x}_i$.

If the error terms are heteroscedastic, proper weighting of the squared residuals will increase the precision of the estimator (cf. weighted linear least squares estimators).

4.4.3 Maximum likelihood v. least squares estimators

Consider following model:

$$y_i = h\left(\boldsymbol{x}_i, \boldsymbol{\theta}_0\right) + \epsilon_i \quad \text{with} \quad i = 1, \cdots, N, \tag{4.28}$$

and let us assume the error terms ϵ_i all to be independent Gaussian random variables having zero-mean and variance σ_i^2. The PDF is then given by:

$$f_\epsilon\left(\epsilon_i\right) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{\epsilon_i^2}{2\sigma_i^2}} \tag{4.29}$$

Substiting ϵ_i by $y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta}_0)$ in Eq. (4.29), results in a PDF for the measured variable y_i:

$$f_y\left(y_i\right) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{\left(y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta}_0)\right)^2}{2\sigma_i^2}}. \tag{4.30}$$

The joint PDF is given by:

$$f_y\left(\boldsymbol{y}\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{\left(y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta}_0)\right)^2}{2\sigma_i^2}}. \tag{4.31}$$

Obviously, the joint likelihood function then becomes

$$\mathcal{L}\left(\boldsymbol{\theta}|\boldsymbol{y}\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{\left(y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta})\right)^2}{2\sigma_i^2}}. \tag{4.32}$$

Maximizing Eq. (4.32) over $\boldsymbol{\theta}$ leads to the same argument as minimizing $-\log \mathcal{L}\left(\boldsymbol{\theta}|\boldsymbol{y}\right)$ over $\boldsymbol{\theta}$. The maximum likelihood estimator can thus be written as:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_N &= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} -\log \mathcal{L}\left(\boldsymbol{\theta}|y_i\right) \\
&= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} -\log \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{\left(y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta})\right)^2}{2\sigma_i^2}} \\
&= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \frac{\log 2\pi}{2} + \log \sigma_i + \frac{\left(y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta})\right)^2}{2\sigma_i^2}
\end{aligned} \tag{4.33}$$

Since the first two terms on the right hand side are not functions of $\boldsymbol{\theta}$ and the fact that the estimator is scale invariant, the estimator can further be simplified:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_N &= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \frac{\left(y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta})\right)_i^2}{\sigma_i^2} \\
&= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} w_{ii} \left(y_i - h(\boldsymbol{x}_i, \boldsymbol{\theta})\right)^2,
\end{aligned} \tag{4.34}$$

with $w_{ii} = \frac{1}{\sigma_i^2}$. The last expression is the same as the weighted least squares estimator. One can thus conclude that for N measurements that are described by a

joint Gaussian distribution, the maximum likelihood estimator is given by a weighed least squares estimator for which the weight terms are the inverse of the variance of respective the measurement. Consequently, the best linear estimator can only be the minimum variance estimator if the error terms are Gaussian distributed.

Bibliography

A. van den Bos. *Parameter Estimation for Scientists and Engineers.* John Wiley and Sons, 2007.

CHAPTER 5

Ordinary least squares estimators

Contents

The work in this chapter has been published as:
J. Veraart, J. Sijbers, S. Sunaert, A. Leemans, and B. Jeurissen. Weighted linear least squares estimation of diffusion MRI parameters: strenghts, limitations and pitfalls. *NeuroImage*, 81:335-346, 2013.

5.1 Introduction

Diffusion magnetic resonance imaging (dMRI) is currently the only method for in vivo and non-invasive quantification of water diffusion in biological tissue [Le Bihan and Johansen-Berg, 2012]. Several diffusion models have been proposed to obtain quantitative diffusion measures, which could provide novel information on the structural and organizational features of biological tissues, the brain white matter (WM) in particular. Typical examples of such diffusion models are diffusion tensor imaging (DTI; [Basser et al., 1994b]) and diffusion kurtosis imaging (DKI; [Jensen et al., 2005]). Both diffusion models have in common that they can be linearized by the natural log-transformation for computing the model parameters. Unsurprisingly, the class of least squares (LS) estimators – both linear and nonlinear – is very popular in diffusion MRI. Indeed, many widely used software packages (e.g., ExploreDTI v4.8.2 [Leemans et al., 2009], FSL v5.0.1 [Jenkinson et al., 2012], Tortoise v1.3.1 [Pierpaoli et al., 2010], Camino [Cook et al., 2006], Slicer v4.2.1 [Pieper et al., 2006]) and, as such, numerous researchers working in the field of dMRI, adopted those least squares estimators. In this chapter, we will introduce and compare different least squares estimators in terms of accuracy and precision. By doing so, we aim to obtain more insight in the strengths, limitations, and potential pitfalls of this simple, though elegant, class of diffusion parameter estimators.

5.2 b-matrix

Both the DTI and DKI model – given by Eq. (2.25) and Eq. (2.35), respectively, can be written as:

$$\boldsymbol{S}(\boldsymbol{B},\boldsymbol{\beta}) = \exp(\boldsymbol{B}\boldsymbol{\beta}), \tag{5.1}$$

or, alternatively

$$\ln \boldsymbol{S}(\boldsymbol{B},\boldsymbol{\beta}) = \boldsymbol{B}\boldsymbol{\beta}, \tag{5.2}$$

with $\boldsymbol{S}(\boldsymbol{B},\boldsymbol{\beta})$ the N model evaluations given $\boldsymbol{B}$, a model-specific design matrix covering all diffusion gradient information, i.e. unit-length diffusion gradient directions $(\{[\hat{g}_{ix},\hat{g}_{iy},\hat{g}_{iz}]^T : i = 1\cdots N\})$ and diffusion strengths $(\{b_i : i = 1\cdots N\})$. The design matrix is typically called the b-matrix [Mattiello et al., 1997]. Furthermore, $\boldsymbol{\beta}$ is the diffusion model's parameter vector, including all independent tensor elements and the noise-free nondiffusion-weighted signal $S_0(\mathbf{0})$. For DTI, $\boldsymbol{\beta}$ is given by:

$$\boldsymbol{\beta} = [\ln S_0(\mathbf{0}), D_{xx}, D_{xy}, D_{xz}, D_{yy}, D_{yz}, D_{zz}]^T, \tag{5.3}$$

whereas for DKI, the parameter vector $\boldsymbol{\beta}$ is given:

$$\begin{aligned}\boldsymbol{\beta} = \Big[&\ln S_0(\mathbf{0}), D_{xx}, D_{xy}, D_{xz}, D_{yy}, D_{yz}, D_{zz}, \tilde{K}_{xxxx}, \tilde{K}_{xxxy}, \cdots \\ &\tilde{K}_{xxxz}, \tilde{K}_{xxyy}, \tilde{K}_{xxyz}, \tilde{K}_{xxzz}, \tilde{K}_{xyyy}, \tilde{K}_{xyyz}, \tilde{K}_{xyzz}, \cdots \\ &\tilde{K}_{xzzz}, \tilde{K}_{yyyy}, \tilde{K}_{yyyz}, \tilde{K}_{yyzz}, \tilde{K}_{yzzz}, \tilde{K}_{zzzz}\Big]^T\end{aligned} \tag{5.4}$$

Although several closely related definitions do exist for the b-matrix, $\boldsymbol{B}$ must be defined as following $N \times 7$ matrix:

$$\boldsymbol{B} = \begin{bmatrix} 1 & -b_1\hat{g}_{1x}^2 & -2b_1\hat{g}_{1x}\hat{g}_{1y} & -2b_1\hat{g}_{1x}\hat{g}_{1z} & -b_1\hat{g}_{1y}^2 & -2b_1\hat{g}_{1y}\hat{g}_{1z} & -b_1\hat{g}_{1z}^2 \\ 1 & -b_2\hat{g}_{2x}^2 & -2b_2\hat{g}_{2x}\hat{g}_{2y} & -2b_2\hat{g}_{2x}\hat{g}_{2z} & -b_2\hat{g}_{2y}^2 & -2b_2\hat{g}_{2y}\hat{g}_{1z} & -b_2\hat{g}_{2z}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -b_N\hat{g}_{Nx}^2 & -2b_N\hat{g}_{Nx}\hat{g}_{Ny} & -2b_N\hat{g}_{Nx}\hat{g}_{Nz} & -b_N\hat{g}_{Ny}^2 & -2b_N\hat{g}_{Ny}\hat{g}_{Nz} & -b_N\hat{g}_{Nz}^2 \end{bmatrix}.$$

for DTI, whereas, for DKI, the b-matrix needs to be extended to an $N \times 22$ matrix. More specifically, the n^{th} row of the $\boldsymbol{B}$ must be complemented with:

$$\frac{b_n^2}{6}\left[\hat{g}_{nx}^4 \quad 4\hat{g}_{nx}^3\hat{g}_{ny} \quad 4\hat{g}_{nx}^3\hat{g}_{nz} \quad 6\hat{g}_{nx}^2\hat{g}_{ny}^2 \quad 12\hat{g}_{nx}^2\hat{g}_{ny}\hat{g}_{nz} \quad 6\hat{g}_{nx}^2\hat{g}_{nz}^2 \quad 4\hat{g}_{nx}\hat{g}_{ny}^3 \quad \cdots \right.$$
$$\left. 12\hat{g}_{nx}\hat{g}_{ny}^2\hat{g}_{nz} \quad 12\hat{g}_{nx}\hat{g}_{ny}\hat{g}_{nz}^2 \quad 4\hat{g}_{nx}\hat{g}_{nz}^3 \quad \hat{g}_{ny}^4 \quad 4\hat{g}_{ny}^3\hat{g}_{nz} \quad 6\hat{g}_{ny}^2\hat{g}_{nz}^2 \quad 4\hat{g}_{ny}g_{nz}^3 \quad \hat{g}_{nz}^4\right].$$

5.3 Non-linear Least Squares estimator

Consider a set of N independently Rice distributed diffusion-weighted signals $\tilde{\boldsymbol{S}}$. Now, $\tilde{\boldsymbol{S}}$ can be modeled as:

$$\tilde{\boldsymbol{S}} = \exp\left(\boldsymbol{B}\boldsymbol{\beta}_0\right) + \boldsymbol{\epsilon}, \tag{5.5}$$

with $\boldsymbol{\beta}_0$ the underlying model parameter vector and $\boldsymbol{\epsilon}$ the column vector with independent error terms. Obviously, the number of independent observations/equations (N) must be greater than or equal to the number of model parameters (7 for DTI and 22 for DKI). If so, the model parameter vector $\boldsymbol{\beta}$ can be estimated by the minimization of the sum of squared deviations about the model predictions:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \| \tilde{\boldsymbol{S}} - \exp\left(\boldsymbol{B}\boldsymbol{\beta}\right) \|_2^2, \tag{5.6}$$

The estimator has no closed-form solution and needs to be solved iteratively until convergence. A typical algorithm to tackle such optimization problems is the Levenberg-Marquardt algorithm. Such iterative optimizers are prone to get stuck in local minima. Therefore, it is important to have a proper initialization. Commonly, (weighted) linear least squares estimators are chosen for this purpose.

5.4 Linear Least Squares estimator

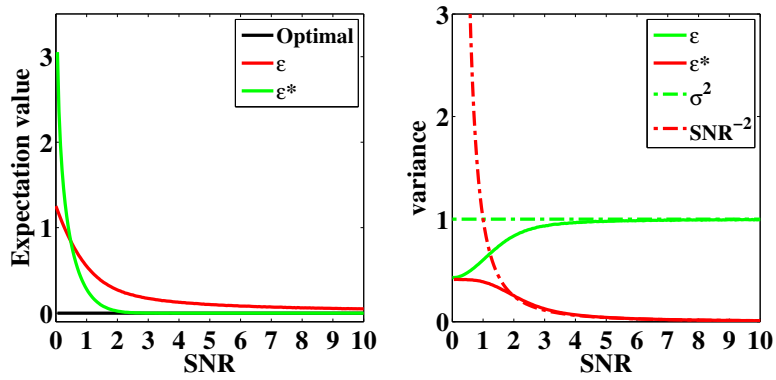
Given the linearized DTI/DKI model:

$$\ln \tilde{\boldsymbol{S}} = \boldsymbol{B}\boldsymbol{\beta}_0 + \boldsymbol{\epsilon}^*, \tag{5.7}$$

the linear least squares (LLS) estimator of $\boldsymbol{\beta}$ is:

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{B}^T \ln \tilde{\boldsymbol{S}}. \tag{5.8}$$

The estimator has a unique closed-form solution if the design matrix has full column rank. It is well recognized that the log-transformed diffusion-weighted signals are



(a) (b)

Figure 5.1: The expectation value (a) and variance (b) of the error term before and after linearization by taking the natural logarithm of the Rice distributed variables is shown as a function of the SNR. In the log-Rice framework, the residuals have zero expectation if SNR > 2. The variance, on the other hand, is then well-approximated by SNR^{-2}. Both curves indicate the potentially high accuracy and precision of WLLS. In the Rice framework, the slow convergence to null expectation reduces the accuracy of the NLS estimator.

heteroscedastic [Basser et al., 1994a, Koay et al., 2006, Salvador et al., 2005]. Therefore, a weighted linear least squares (WLLS) approach with well-defined weights, i.e. the inverse of the log-transformed signals' variances, is expected to provide more precise diffusion parameter estimates, at least, compared to its unweighted linear alternative.

5.5 Weighted Linear Least Squares estimator

Salvador et al. [2005] also showed that the variance of $\boldsymbol{\epsilon}^*$ depends on the respective noise-free signals (see Fig. 5.1b):

$$Var(\boldsymbol{\epsilon}^*) \approx \left[\frac{\sigma_c^2}{S_0^2(\boldsymbol{B}_{1*})}, ..., \frac{\sigma_c^2}{S_0^2(\boldsymbol{B}_{N*})}\right], \tag{5.9}$$

with $\boldsymbol{S}_0(\boldsymbol{B}) = \{S_0(\boldsymbol{B}_{i*}) : i = 1, ..., N\}$ the noise-free diffusion-weighted signals and σ_c the noise level if SNR exceeds two for each datum. The best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$, i.e. the estimator with the highest precision within the class of unbiased linear estimators, can only be designed by including weights that equal the (scaled) reciprocal of the variance of the corresponding error terms (Eq. (5.9)):

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{B}^T\boldsymbol{W}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^T\boldsymbol{W}\ln\tilde{\boldsymbol{S}}, \tag{5.10}$$

with

$$\boldsymbol{W} = \mathrm{diag}\left(\boldsymbol{S}_0^2(\boldsymbol{B})\right). \tag{5.11}$$

Obviously, the noise-free diffusion-weighted signals are not known and, as such, the weight matrix $\boldsymbol{W}$ needs to be estimated. Different ways to approximate the theoretically optimal weights have been presented [Basser et al., 1994b, Salvador et al., 2005] and adopted by the community. However, most often, information on the weighting is not provided in scientific reports or software documentation. Nonetheless, suggested approaches are:

(a) WLLS_1: $\tilde{\boldsymbol{W}} = \text{diag}\left(\tilde{\boldsymbol{S}}^2\right)$, the weight are the squares of the respective noisy diffusion-weighted signals [Basser et al., 1994a, Koay et al., 2006];

(b) WLLS_2: $\tilde{\boldsymbol{W}} = \text{diag}\left(\exp\left(2\boldsymbol{B}\hat{\boldsymbol{\beta}}_{\text{LLS}}\right)\right)$, the weights are the preliminary estimates of $\boldsymbol{S}^2(b, \boldsymbol{g})$, reconstructed from the LLS estimate of $\boldsymbol{\beta}$ [Salvador et al., 2005].

The subindices of WLLS indicate that (a) and (b) are single-step and dual-step strategies, respectively. The estimation of $\boldsymbol{W}$ can iteratively be improved [Salvador et al., 2005]. The weight matrix for the n^{th} iteration is given by the predicted diffusion-weighted signals from the previous estimate of the diffusion model parameters ($\hat{\boldsymbol{\beta}}_{n-1}$):

$$\tilde{\boldsymbol{W}}_n = \text{diag}\left(\exp\left(2\boldsymbol{B}\hat{\boldsymbol{\beta}}_{n-1}\right)\right). \tag{5.12}$$

This iterative WLLS can be initialized by (a) or (b), which will be referred to as IWLLS_1 and IWLLS_2, respectively.

5.6 Strengths, limitations, and pitfalls

5.6.1 General comment

Let's start with a shot from the hip: a drop in accuracy is the price to pay when using ordinary least squares estimators in diffusion MRI. The SNR dependent difference between the expectation value of a Rice – or more generally, a noncentral χ – distributed value and its underlying noise-free value causes inaccuracies in the estimation of the diffusional tensors. Since the SNR of the dMRI signals depends on factors such as diffusion gradient direction, b-value, diffusivity, and diffusion anisotropy, the so-called noise bias is omnipresent in quantitative dMRI. Indeed, many *noise artifacts* have been discussed earlier. For example, the (directional) diffusivity coefficients will be underestimated. The lower the SNR of the diffusion-weighted signals, the stronger the underestimation will be. Since a low SNR can result from a high b-value and/or a high direction diffusivity, one can immediately conclude that the noise bias will be more expressed at high b-values or along the gradient direction with the highest diffusivity. Indeed, axial diffusivity will more strongly be underestimated compared to radial diffusivity. Therefore, an underestimation in FA is a direct consequence (see Fig. 5.2). This phenomenon is often referred to as *squashing the ADC*[1] *peanut* [Jones and Basser, 2004]. However, for low anisotropy factors, overestimation of FA was observed previously [Pierpaoli

[1] ADC: apparent diffusion coefficient

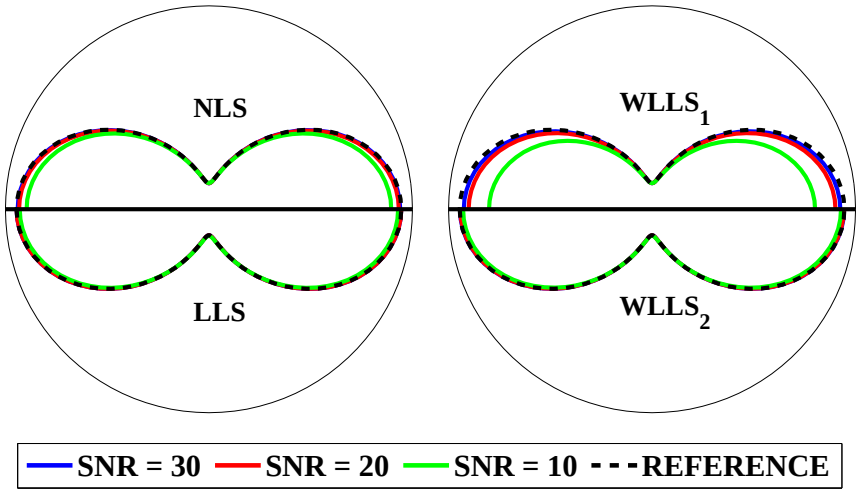


Figure 5.2: Average ADC profiles estimated using different least squares estimators. The *squashing the ADC peanut* strongly depend on the used estimator. Theoretically, the (weighted) linear least squares estimators have the potential to be more accurate than the nonlinear least squares estimator. However, its actual accuracy will depend on the selected weight terms. The SNR is here defined as the ratio of the nondiffusion-weighted signal and the noise level. The SNR of the diffusion-weighted images are inherently lower and will depend on the gradient direction.

and Basser, 1996]. The increase in FA originates in the eigenvalue repulsion, which is a general noise effect rather than a consequence of the elevated expectation value of the noncentral χ distribution. The b-value dependency of the noise bias, on its turn, lead to an overestimation of the directional kurtosis coefficients [Veraart et al., 2011]. Even a quick look at Fig. 5.2 already suggests the importance of a well-considered choice of the diffusion parameter estimator. Indeed, in terms of accuracy, the different estimators shows different behavior. At first glance, the LLS appears more accurate in the estimation of the diffusion tensor, compared to NLS. Furthermore, the weighting strategy in case of WLLS highly impact the estimators' performance. After some theoretical considerations, simulation and real data experiments are done to evaluate the different estimation strategies.

5.6.2 Theoretical considerations

Salvador et al. [2005] derived an analytical expression for the expectation value of the error term, given that the linearized diffusion model is fitted to the natural logarithm of Rice distributed diffusion-weighted measurements ($\tilde{\boldsymbol{S}}$). The probability distribution function (PDF) of an arbitrary Rice distributed variable m is given by Sijbers et al. [1998]:

$$f_m\left(m|m_0,\sigma\right) = \frac{m}{\sigma^2}\exp\left(-\frac{m^2+m_0^2}{2\sigma^2}\right) I_0\left(\frac{m\,m_0}{\sigma^2}\right), \tag{5.13}$$

with m_0 the noise-free signal, σ^2 the noise variance, and I_0 the zeroth order modified Bessel function of the first kind. The PDF of the log-transformed variable $m^* = \ln(m)$ can be determined by substituting m in Eq. (5.13) by m^*:

$$\begin{aligned} f_{m^*}(m^*|m_0,\sigma) &= f_m(\exp(m^*)|m_0,\sigma)\,\frac{d\exp(m^*)}{dm^*} \\ &= \frac{e^{2m^*}}{\sigma^2}\exp\left(-\frac{e^{2m^*}+m_0^2}{2\sigma^2}\right) I_0\left(\frac{e^{m^*}m_0}{\sigma^2}\right). \end{aligned} \tag{5.14}$$

The expectation value of m^* is given by:

$$\begin{aligned} \mathbb{E}[m^*] &= \int_{-\infty}^{\infty} m^* f_{m^*}(m^*|m_0,\sigma)\,dm^* \\ &= \int_{\mathrm{SNR}^2/2}^{\infty} \frac{1}{t\,e^t}dt + \ln(m_0), \end{aligned} \tag{5.15}$$

with $\mathrm{SNR} = \frac{m_0}{\sigma}$. If the error term ϵ^* is defined as $m^* - \ln(m_0)$, then its expectation value can immediately be derived:

$$\begin{aligned} \mathbb{E}[\epsilon^*] &= \mathbb{E}[m^* - \ln(m_0)] \\ &= \int_{\mathrm{SNR}^2/2}^{\infty} \frac{1}{t\,e^t}dt + \ln(m_0) - \ln(m_0) \\ &= \int_{\mathrm{SNR}^2/2}^{\infty} \frac{1}{t\,e^t}dt. \end{aligned} \tag{5.16}$$

In Fig. 5.1a, $\mathbb{E}[\epsilon^*]$ is shown for different SNR values. It can be seen that for $\mathrm{SNR} > 2$, the error term is zero-centered. The expectation value of $\epsilon = m - m_0$ is given by:

$$\mathbb{E}[\epsilon] = \sigma_c\sqrt{\frac{\pi}{2}}L_{1/2}\left(-\frac{m_0^2}{2\sigma_c^2}\right) - m_0. \tag{5.17}$$

In Fig. 5.1a, we also showed $\mathbb{E}[\epsilon]$ as a function of the SNR. The expectation value of the error terms is always larger than zero. Based on the comparison of both curves, one might expect only the LLS to be unbiased, at least, if $\mathrm{SNR} > 2$.

The SNR dependency of the variance of ϵ^* cannot be ignored. To increase the precision of the linear estimators, weights needs to be added to the LLS. However, adding weights might lower the accuracy of the estimator. Indeed, the expectation value of the error term of the WLLS, ϵ_w^*, might be larger than zero, even with $\mathrm{SNR} > 2$:

$$\begin{aligned} \mathbb{E}[\epsilon_w^*] &= \mathbb{E}[w\epsilon^*] \\ &= \mathbb{E}[w]\,\mathbb{E}[\epsilon^*] + \sigma_w\sigma_{\epsilon^*}\mathrm{corr}(w,\epsilon^*) \\ &= \sigma_w\sigma_{\epsilon^*}\mathrm{corr}(w,\epsilon^*), \end{aligned} \tag{5.18}$$

with σ_w and σ_{ϵ^*} the standard deviation of the weights and the unweighted error term, respectively. Note that the last step only holds if $\mathrm{SNR} > 2$, because then

$\mathbb{E}[\epsilon^*] = 0$. If w is deterministic or uncorrelated to ϵ^*, then the WLLS will also be unbiased. This, however, cannot be guaranteed in practice. Indeed, often the weights are the diffusion-weighted signals themselves (cf. WLLS_1). Due to the high correlation with the unweighted error term, a bias will be introduced by weighting the linear least squares estimator. Attempts to lower the σ_w or $\text{corr}(w, \epsilon^*)$ are expected to give improved accuracy, at least compared to WLLS_1. Therefore, one might expect WLLS_2 to be more accurate than WLLS_1.

5.6.3 Experimental validation

5.6.3.1 Single voxel simulations

Monte Carlo simulations (50000 trials) were performed to evaluate the performance of the different strategies in the estimation of FA, MD, and MK. In addition to the linear estimators, the ordinary nonlinear least squares (NLS) estimator, initialized by a guess through WLLS_2, was evaluated. Throughout all experiments, FA and MD were estimated with the DTI model, whereas the MK was estimated with the DKI model. First, the accuracy and the precision of the different estimators were evaluated as a function of the SNR of the Rician distributed nondiffusion-weighted signals, that is, SNR_{b_0}. Indeed, SNR_{b_0} is here defined as the ratio between the nondiffusion-weighted signal and the noise level (cf. Jones and Basser [2004]). Note that the SNR of the diffusion-weighted signals are always lower because of the diffusion-dependent signal attenuation. For the DTI model, the MD and FA were set to $0.8 \times 10^{-3}\,\text{mm}^2/\text{s}$ and 0.85, respectively. The b-value was $1000\,\text{s/mm}^2$, and 60 gradient directions – isotropically distributed over a unit sphere using Coulomb's law of repulsion [Jones et al., 1999] – were used. We included five nondiffusion-weighted signals. For the DKI simulations, additional diffusion-weighted signals with $b = 2500\,\text{s/mm}^2$ were sampled along the same 60 directions. The MK was defined as 1.05, which is in agreement with values typically observed in the corpus callosum of the healthy human brain [Lätt et al., 2013]. Noisy synthetic data were obtained by adding zero-mean complex Gaussian noise to the noise-free diffusion-weighted signals, which were calculated from the ground truth tensors using Eq. (5.1). The absolute value of the resulting complex noisy signals was taken afterwards to obtain their magnitudes. First, in Fig. 5.3, the different estimators are compared in terms of accuracy, precision, and MSE as a function of $\text{SNR}{-}1_{b_0}$. The FA, MD, and MK values – calculated from the average model parameters to exclude nonlinear effects such as eigenvalue repulsion [Pierpaoli and Basser, 1996]– strongly vary across the different estimators (see Fig. 5.3(a-c)). Generally, the NLS estimator shows a large difference to the reference values, compared to the unweighted and weighted linear approaches. However, note that non-optimal weighting strongly reduce the accuracy of the estimator. The practical WLLS approaches, i.e. WLLS_1 and WLLS_2, show a lower accuracy than the LLS and BLUE. Remarkably, the drop in accuracy is especially large for the WLLS_1. While the accuracy of the WLLS_2 is only slightly lower than BLUE, the WLLS_1 performs even worse than the NLS in terms of accuracy. At very low SNR, e.g., due to the use of high b-values, all least squares estimators are inherently biased. In terms of precision (see Fig. 5.3(d-f)), all weighted linear estimators outperform the

unweighted alternative. The LLS has low performance in terms of MSE because of the low precision, whereas the high MSE of the $WLLS_1$ results from its low accuracy (see Fig. 5.3(g,h)). For the MSE in the estimation of MK (see Fig. 5.3(i)), the low accuracy of the NLS and the $WLLS_1$ is strongly counterbalanced by their high precision. Therefore, both estimators have a relatively low MSE. Simulations beyond these single-voxel experiments are needed to avoid overinterpretation of that observation. In general, the differences between the estimators diminish with increasing SNR_{b_0}. The same could be observed by lowering b-value or simulated diffusion coefficients. Second, we evaluated the influence of the number of gradient directions per b-value shell on the accuracy of WLLS. Unlike the initial simulation experiment, the SNR_{b_0} was kept constant at the level of 20, whereas the number of gradient direction per shell varied from 6 to 120 for DTI and from 15 to 120 for DKI. In Fig. 5.4, the influence of the number of gradient directions per b-value shell on the accuracy of the different estimators is shown. The accuracy of the $WLLS_2$ increases steadily with increasing number of gradient directions due to the increasing precision of the LLS estimator or, as such, the increasing precision of the predicted diffusion-weighted signals. By contrast, with a low number of gradient directions, the difference between the $WLLS_1$ and the $WLLS_2$ in terms of accuracy vanishes due the reduced performance of the $WLLS_2$. Third, the effect of increasing the number of iterations (n) on the iterative WLLS approaches was evaluated. Now, the SNR_{b_0} was kept constant at the level of 20, whereas n varied from one to ten. In Fig. 5.5, the effect of iterations on the performance of the (I)WLLS estimators is shown. The graphs indicate that after a few iterations, both IWLLS estimators already closely approximate the performance of the BLUE, in terms of accuracy, precision, and MSE. The effect of the initial weighting matrix already vanishes for $n > 1$. The same findings are observed for all diffusion metrics.

5.6.3.2 Whole brain simulations

Rice distributed simulation data sets, representing the whole human brain white matter, rather than a single voxel, were used for comparing the MSE of the diffusion parameters obtained by the different least squares estimators. The simulated data sets were constructed as follows. First, ground truth tensors were obtained by voxel-wise fitting the DKI model to a real data set (see section 5.6.3.3: Data set 1). Second, a set of noise-free diffusion-weighted signals was reconstructed from those diffusion tensors using the DTI and DKI model for the DTI and DKI analyses, respectively. Gradient directions and b-values were in agreement with the single-voxel experiments. Third, 5000 sets of noisy diffusion-weighted signals with a uniform SNR_{b_0} of 20 were generated by adding 5000 realizations of complex Gaussian noise to the noise-free diffusion-weighted signals. From each set, the diffusion tensors were estimated. From the 5000 trials, the accuracy and precision of the different estimators were evaluated. In Fig. 5.6, scatter plots show the relationship between the accuracy, precision and MSE in the estimation of the diffusion model parameters of $WLLS_1$ and $WLLS_2$ against that of LLS. Each point in the scatter plot corresponds to a single voxel of the simulated data set. In Fig. 5.7, $WLLS_1$ and $WLLS_2$ are compared to NLS, whereas in Fig. 5.8,

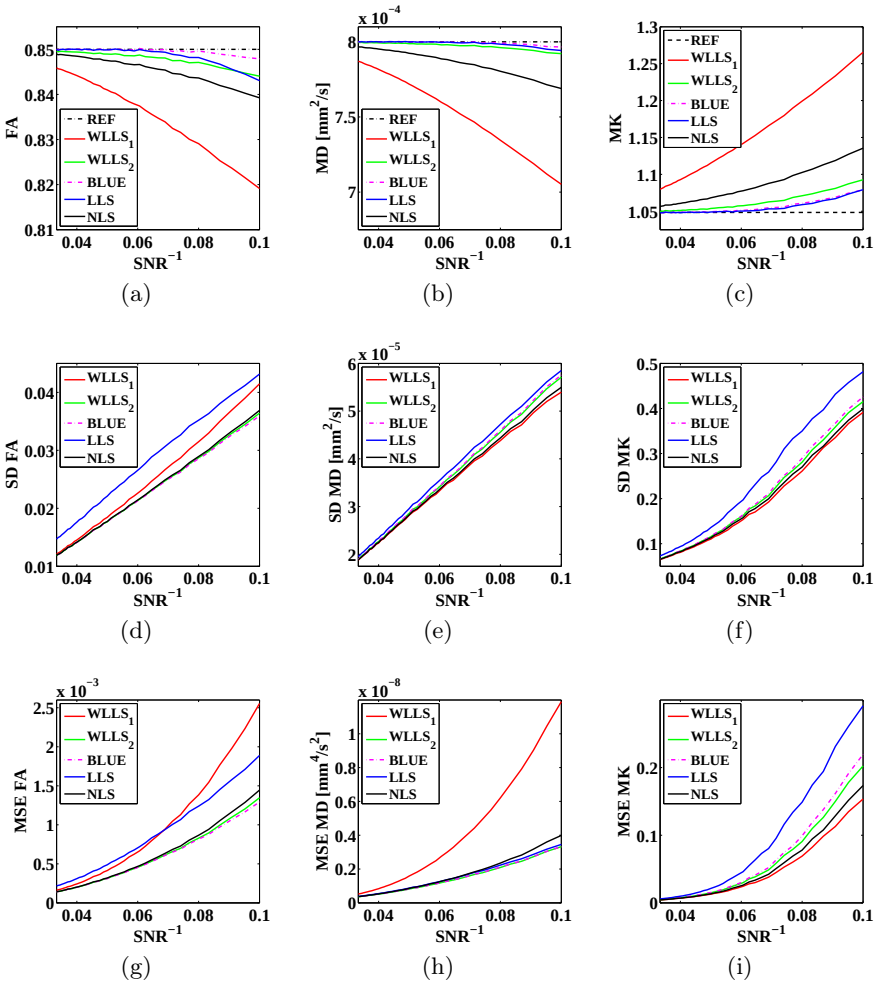


Figure 5.3: Simulation experiment 1: FA, MD, and MK calculated from the average model parameters (a-c), the standard deviation (SD) of the estimated diffusion parameters (d-f), and MSE in the estimation of FA, MD, and MK (g-i) are shown as a function of the $\mathrm{SNR}_{b_0}^{-1}$ for the different least squares estimators.

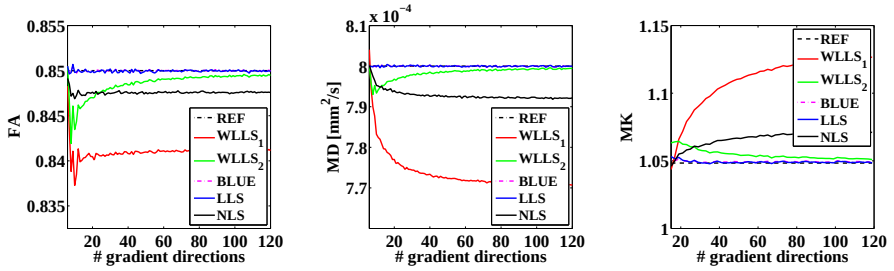


Figure 5.4: Simulation experiment 2: FA, MD, and MK calculated from the averaged model parameters are shown as a function of the number of gradient directions per b-value shell for the different least squares estimators.

IWLLS_1 and IWLLS_2 are compared to their non-iterated counterparts. First, on average, the WLLS_1 has a significantly large bias in the estimation of the diffusion model parameters, compared to the LLS, WLLS_2, $\text{IWLLS}_{1,2}$ ($n = 5$), and NLS. Indeed, WLLS_1 significantly underestimates FA and MD. The underestimation is significantly larger than the underestimation of FA and MD observed for the NLS. The NLS, on its turn, is significantly outperformed by the multi-step WLLS approaches and the LLS in terms of accuracy in the estimation of the diffusion tensor. Moreover, the one-sample Wilcoxon signed rank test demonstrated that LLS, WLLS_2, $\text{IWLLS}_{1,2}$ are not significantly biased. For MK, similar conclusion can be drawn. However, this time, all least squares estimators show a significant overestimation of the MK. Again, LLS, WLLS_2, and $\text{WLLS}_{1,2}$ show significantly higher accuracy compared to NLS and, especially WLLS_1. Second, in terms of precision, LLS is significantly outperformed by all of its weighted alternatives as theoretically expected. In general, NLS showed higher precision in the estimation of all diffusion parameters than the WLLS approaches. However, in the estimation of MD and MK, the WLLS_1 showed the highest precision. Third, on average, WLLS_1 showed the highest MSE in the estimation of MD and MK in comparison to all other estimators. The estimator with the least average MSE in the estimation of MD was the WLLS_2. For MK, WLLS_2 was significantly outperformed by NLS, which compensated its low accuracy by its high precision. The NLS also showed the least MSE in the estimation of FA. The MSE in the estimation of FA, MD and MK was significantly lower for WLLS_2 in comparison to IWLLS_2, whereas IWLLS_1 had always significantly lower MSE than its non-iterated alternative. During these simulations, statistical significance ($p < 0.01$) was always shown with a paired Wilcoxon signed rank test.

5.6.3.3 Real data experiments

The following diffusion-weighted data sets of different healthy volunteers were acquired:

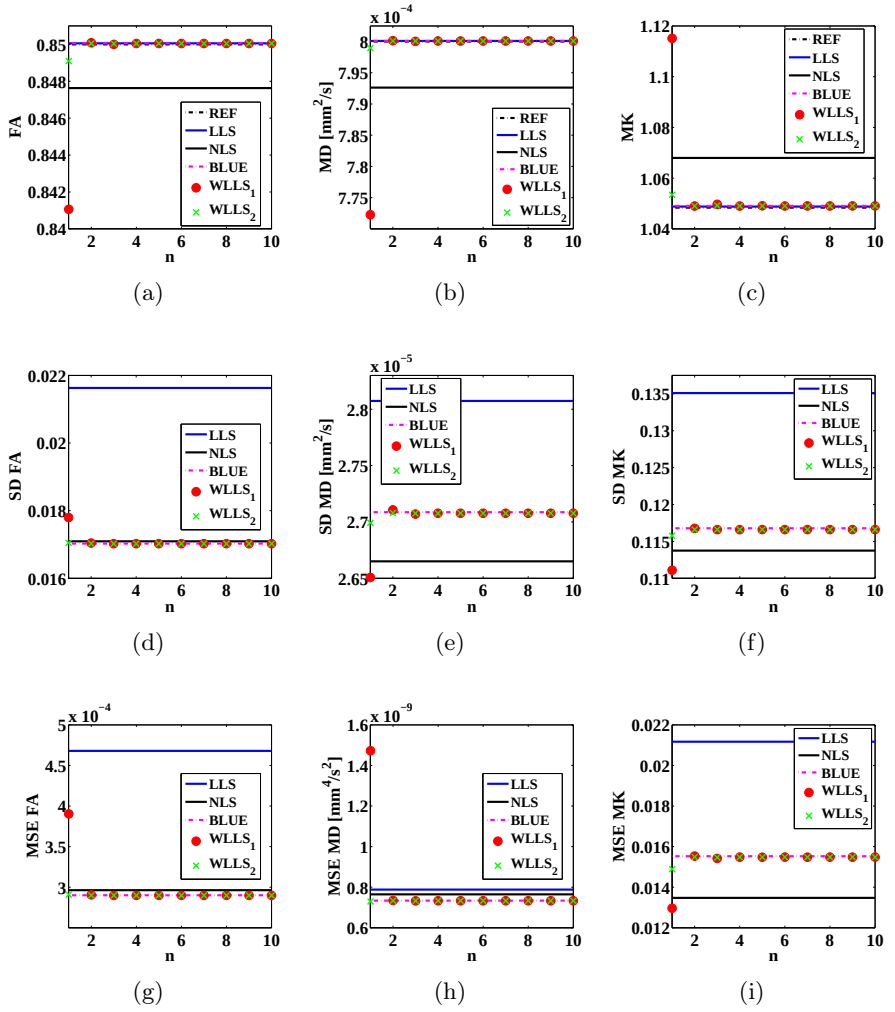
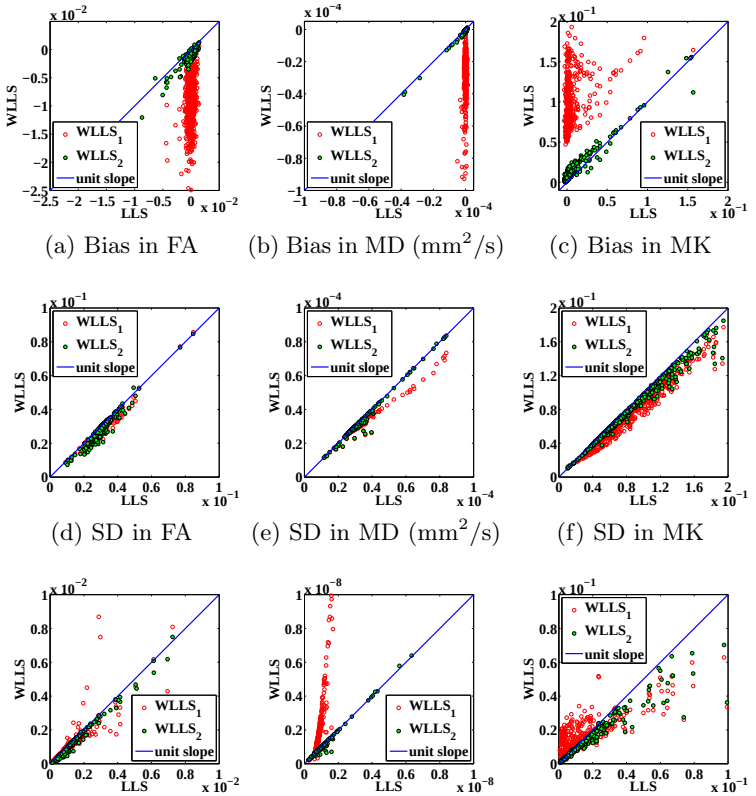


Figure 5.5: Simulation experiment 3: FA, MD, and MK calculated from the average model parameters (a-c), the standard deviation of the estimated diffusion parameters (d-f), and MSE in the estimation of FA, MD, and MK (g-i) are shown as a function of the number of iterations (n) for the different least squares estimators.

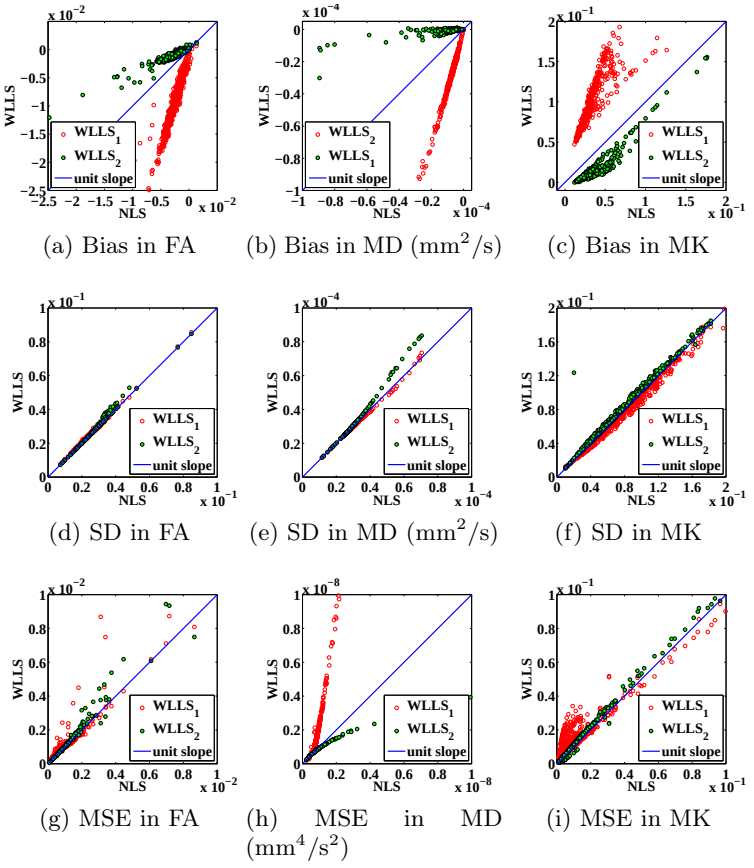


(a) Bias in FA (b) Bias in MD (mm^2/s) (c) Bias in MK

(d) SD in FA (e) SD in MD (mm^2/s) (f) SD in MK

(g) MSE in FA (h) MSE in MD (mm^4/s^2) (i) MSE in MK

Figure 5.6: Simulation experiment 4: Scatter plots show the relationship between the bias, standard deviation and MSE in the estimation of the diffusion model parameters of $WLLS_1$ and $WLLS_2$ against that of LLS. *red circles:* Initial weights of the WLLS approaches are the squares of the respective noisy diffusion-weighted signals – *green dots:* Initial weights are the predicted signal from an estimate of the diffusion model parameters, obtained with the unweighted LLS estimator. The blue lines are unit-slope-lines.

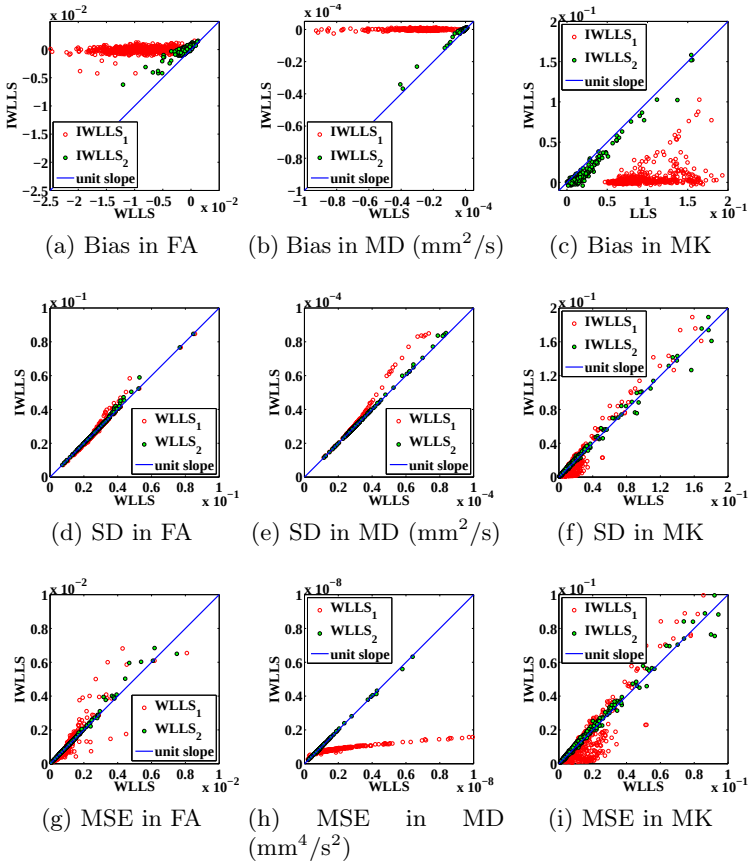


(a) Bias in FA (b) Bias in MD (mm^2/s) (c) Bias in MK

(d) SD in FA (e) SD in MD (mm^2/s) (f) SD in MK

(g) MSE in FA (h) MSE in MD (mm^4/s^2) (i) MSE in MK

Figure 5.7: Scatter plots show the relationship between the bias, standard deviation and MSE in the estimation of the diffusion model parameters of $WLLS_1$ and $WLLS_2$ against that of NLS. *red circles:* Initial weights of the WLLS approaches are the squares of the respective noisy diffusion-weighted signals – *green dots:* Initial weights are the predicted signal from an estimate of the diffusion model parameters, obtained with the unweighted LLS estimator. The blue lines are unit-slope-lines..



(a) Bias in FA (b) Bias in MD (mm^2/s) (c) Bias in MK

(d) SD in FA (e) SD in MD (mm^2/s) (f) SD in MK

(g) MSE in FA (h) MSE in MD (mm^4/s^2) (i) MSE in MK

Figure 5.8: Scatter plots show the relationship between the bias, standard deviation and MSE in the estimation of the diffusion model parameters of $IWLLS_1$ and $IWLLS_2$ against that of their respective non-iterated counterparts. *red circles:* Initial weights of the WLLS approaches are the squares of the respective noisy diffusion-weighted signals – *green dots:* Initial weights are the predicted signal from an estimate of the diffusion model parameters, obtained with the unweighted LLS estimator. The blue lines are unit-slope-lines.

Data set 1: A first diffusion-weighted data set was collected on a 3T Philips Achieva MR scanner, using a 8-channel receiver head coil. Diffusion sensitizing was applied along 60 isotropically distributed gradient directions with $b = 1200\,\mathrm{s/mm^2}$ as well as $b = 2500\,\mathrm{s/mm^2}$. Additionally, one image without diffusion sensitization was acquired. Other imaging parameters were: TR/TE : 10265/107 ms; in-plane resolution: $1.75 \times 1.75\,\mathrm{mm^2}$; NEX: 1; slice thickness: 2 mm; axial slices: 70; and parallel imaging: SENSE with acceleration factor 2.

Data set 2: A second diffusion-weighted data set was acquired on a 1.5T Siemens MR system using a single receiver coil [Leemans et al., 2006]. A gradient configuration with 60 isotropically distributed gradient directions with $b = 700\,\mathrm{s/mm^2}$ was used. 10 nondiffusion-weighted images were additionally acquired. Other acquisition parameters were as follows: TR/TE : 8300/108 ms; in-plane resolution: $2 \times 2\,\mathrm{mm^2}$; NEX: 1; slice thickness: 2 mm; and axial slices: 60.

Data set 3: A third diffusion-weighted data set was acquired on a Siemens Trio (3T) MR scanner, using a 12-channel receiver head coil. Diffusion weighting was applied along 60 isotropically distributed gradient directions with $b = 1000\,\mathrm{s/mm^2}$ as well as $b = 2500\,\mathrm{s/mm^2}$. Additionally, 10 nondiffusion-weighted images were acquired. Other imaging parameters were: TR/TE : 6100/118 ms, in-plane resolution: $2.5 \times 2.5\,\mathrm{mm^2}$; NEX: 1; slice thickness: 2.5 mm; axial slices: 40; and parallel imaging: mSENSE with acceleration factor 2.

The diffusion-weighted data were corrected for motion and eddy currents, including signal modulation and b-matrix rotation [Leemans et al., 2009, Leemans and Jones, 2009]. Next, if the diffusion protocol met the minimal DKI requirements [Lu et al., 2006], a DKI analysis was performed in addition to a DTI analysis. During the DTI analyses, all high b-valued – $b > 1500\,\mathrm{mm^2/s}$ – diffusion-weighted images were excluded. Since *data set 2* has only one nonzero b-value shell, a DKI analysis was not possible. FA and MD maps were calculated from the DTI analysis with the optional DKI analysis also providing a MK map. The percentage differences in the estimation of the diffusion parameters of the different estimators, compared to a *bronze standard* were calculated. As a bronze standard, we adopted the previously proposed parameter estimation framework, consisting of the estimation of a (spatially varying) noise map (see Chapter 7) and an accurate parameter estimator, i.e. the conditional least squares (CLS) estimator, which properly accounts for the Rice distribution of all acquired diffusion-weighted data samples (see Chapter 6). In Fig. 5.9, Fig. 5.10, and Fig. 5.11, the percentage differences in the estimation of FA, MD, and MK of the different estimators, compared to the *bronze standard*, are shown for a single slice of data set 1, 2, and 3, respectively. For each data set, an axial slice composed out of white matter, (deep) gray matter, and cerebrospinal fluid was selected to represent the anatomy of the human brain. The *bronze standard* maps are shown for anatomical reference (left columns). In general, the results are in-line with the simulation results. Indeed, the WLLS_1 and the NLS significantly underestimate the diffusion tensor model parameters in

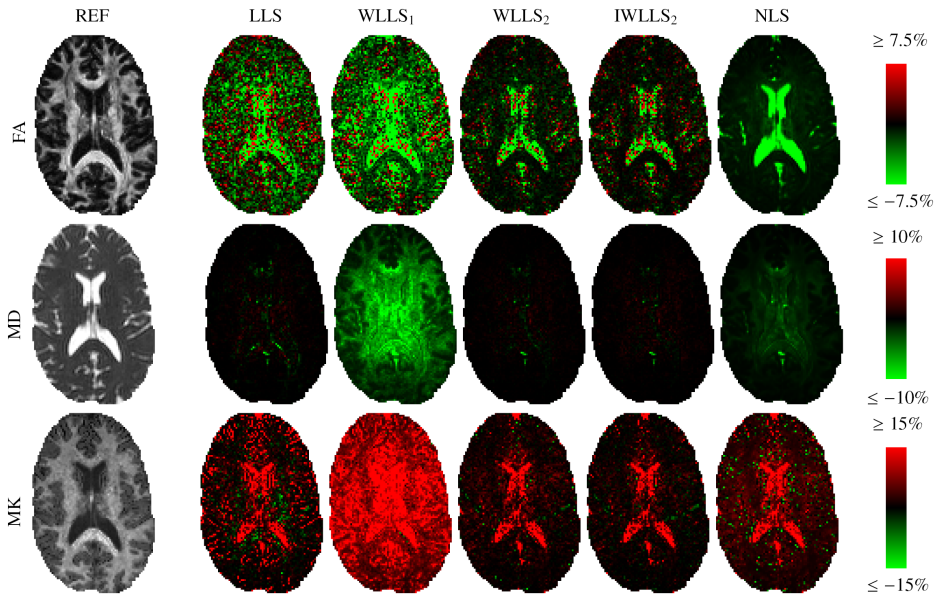


Figure 5.9: Real data set 1: The percentage differences between the FA, MD, and MK estimates for the least squares estimators and the bronze standard, i.e. an accurate diffusion parameter estimation framework, which accounts for the Rice data statistics. The bronze standard diffusion parameter maps are shown for anatomical reference (left column)

both gray and white matter. In all cases, the error is significantly larger for the WLLS_1. In general, the LLS, WLLS_2, and $\text{IWLLS}_{1,2}$ show small underestimation of FA and MD. However, it must be noted that for data set 1, a minimal, though significant, overestimation of MD was observed. This observation might indicate a slight underestimation of the spatially varying noise levels in the bronze standard. Unsurprisingly, for all estimators, the overestimation of MK is significant due to the low SNR of the high b-valued diffusion-weighted images. Again, the overestimation is significantly larger for the WLLS_1, compared to all others. The NLS showed significantly higher error in the estimation of MK than the LLS, WLLS_1, and $\text{IWLLS}_{1,2}$. The paired Wilcoxon signed rank test was always applied to evaluate the pairwise difference between two estimators. The statistical tests included all white and/or gray matter voxels.

5.7 Discussion

The DTI and DKI model are both log-linear models. Hence, the unknown model parameters can be estimated with an LLS estimator, or its weighted variants, after log-transforming the diffusion-weighted MR signals. Those estimators are widely used in DTI and DKI studies, because they come with several strengths. First, the (weighted) LLS estimators have a closed-form solution. Therefore,

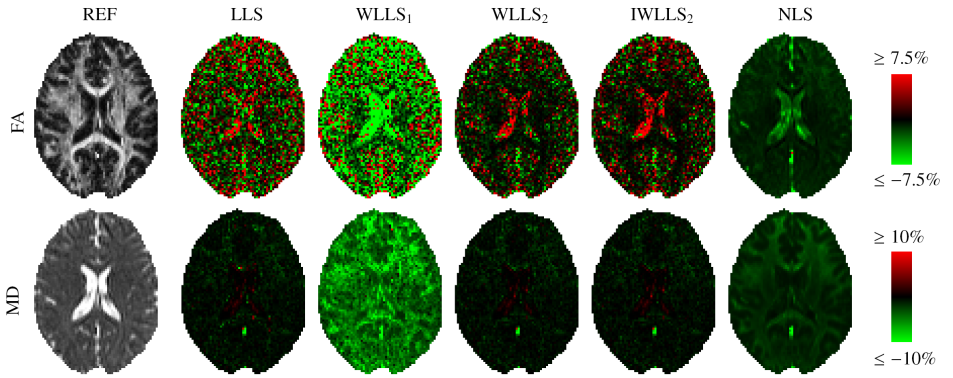


Figure 5.10: Real data set 2: The percentage differences between the FA, MD, and MK estimates for the least squares estimators and the bronze standard.

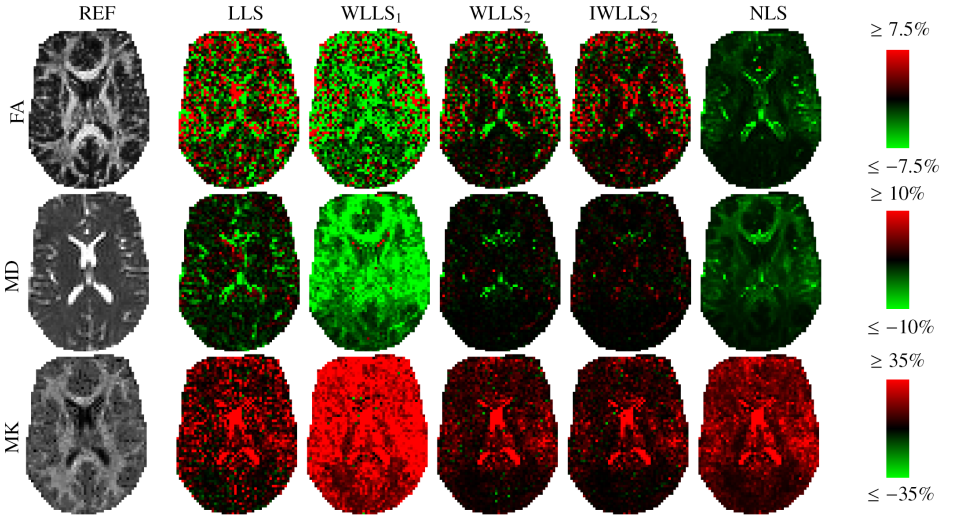


Figure 5.11: Real data set 3: The percentage differences between the FA, MD, and MK estimates for the least squares estimators and the bronze standard.

unlike iterative nonlinear strategies, the linear estimators are computationally efficient and not prone to getting stuck in a local optimum. Second, the linear estimators are potentially very accurate, especially compared to the NLS estimator [Salvador et al., 2005]. Under conditions that are discussed below, the linear estimators are even unbiased due to the zero expectation of the error term in the log-Rice framework. Third, the high accuracy is not at the expense of the ease of use. In other words, unlike many advanced diffusion parameter estimators, e.g., the maximum likelihood (ML) estimator, the linear estimators don't require the knowledge of the noise parameter [Veraart et al., 2012]. The estimation of the noise parameter is not only challenging, it also reduces the precision of the diffusion parameter estimator. Unfortunately, those advantages go hand in hand with some limitations and potential pitfalls. Some of them are to the best of our knowledge still unrecognized.

The unbiasedness of the linear estimators, even under inequality of variances, is subject to two conditions: the SNR of the diffusion-weighted signals should not be too low (> 2) and the diffusion-weighted data are assumed to be Rice distributed before the log-transformation. Those conditions might not be fulfilled because the use of high b-values or high spatial resolution, magnitude image operations prior to model fitting [Rohde et al., 2004, Veraart et al., 2012], or the use of parallel MR imaging techniques [Aja-Fernández et al., 2011, Aja-Fernández and Tristán-Vega, 2012]. Indeed, our simulation and real data experiments confirmed that systematic errors in the calculation of quantitative parameters of clinical interest such as FA, MD, and MK will appear by lowering the SNR. The bias becomes even more prominent for multichannel and/or accelerated MRI reconstruction techniques that generate non-Rice distributed data (e.g., GRAPPA [Griswold et al., 2002] or homodyne partial Fourier reconstruction [Noll et al., 1991]) [Tristán-Vega et al., 2011, Veraart et al., 2012]. Furthermore, the accuracy of the linear estimators drops by applying operations such as motion correction and smoothing on the magnitude diffusion-weighted data as these might change the native data distribution [Veraart et al., 2012]. The mathematical reasoning of Salvador et al. [2005] then no longer holds. However, although motion and eddy current distortion correction was applied prior to tensor fitting in our real data study, the linear approaches still outperformed the NLS estimator in terms of accuracy.

The linearization of both diffusion models comes with the cost of a reduced precision of the diffusion parameter estimators. The cost, however, can be limited by accounting for the heteroscedasticity of the log-transformed data. The variances of the log-transformed Rice distributed magnitude MR signals must be known to design a WLLS estimator with optimal precision. Unfortunately, the variances depend on the unknown noise-free diffusion-weighted signals [Salvador et al., 2005]. Therefore, the best linear unbiased estimator (BLUE), i.e. the unbiased linear estimator with the highest precision, does not exist in practice. Nevertheless, we showed that optimal linear performance can be well approximated by a WLLS estimator for which the weight terms are estimated from noisy data with $\mathrm{SNR} > 2$. The approximation improves by drawing more diffusion-weighted samples or by iteratively updating the estimate of the weight matrix. In practice, only a few iterations are needed for convergence.

Already in the early days of DTI, [Basser et al., 1994a] stated that the weights

for the WLLS can simply be the squares of the respective noisy diffusion-weighted signals (cf. WLLS_1). More than a decade later, the statement was repeated by Koay et al. [2006]. Many widely used software packages (e.g., ExploreDTI v4.8.2 [Leemans et al., 2009], FSL v5.0.1 [Jenkinson et al., 2012], Tortoise v1.3.1 [Pierpaoli et al., 2010]) and, as such, numerous researchers working in the field of dMRI, adopted the approach. In the meantime, Salvador et al. [2005] proposed alternative, multi-step weighting schemes. In those approaches, the weights are the squares of the predicted signals, which are reconstructed from a previous estimate of the diffusion model parameters, obtained with the unweighted LLS estimator (cf. WLLS_2) or a previous iteration of the WLLS estimator (cf. IWLLS). The latter strategy was adopted by software packages such as Camino [Cook et al., 2006] and Slicer v4.2.1 [Pieper et al., 2006]. So, nowadays, different weighting strategies are commonly used in a daily practice. This lack of unity not only obstructs multi-center research, it might also lead to misleading conclusions or irreproducible results. Indeed, we showed in this work that the performance of the WLLS strongly depends on the selected weight terms. More specifically, ill-chosen weights strongly reduce the accuracy of the linear estimator. Simulation and real data experiments indicated that the strategy proposed by Basser et al. [1994a] (WLLS_1) is the least favorable weighting strategy. Therefore, we suggest always opting for multi-step weighting strategies.

In this study, we only reported FA, MD, and MK. Nevertheless, it might be expected that similar conclusions can be drawn for the other tensor-derived metrics to a greater or lesser degree depending on the directionality of the SNR. However, note that unlike diffusion metrics such as axial and radial diffusivity, or axial and radial kurtosis, the principal diffusion direction, i.e. the first eigenvector of the diffusion tensor, is not subject to the Rician noise bias due to the symmetry of the diffusion tensor.

If the number of diffusion-weighted samples and the SNR are low, then the diffusion-weighted signals predicted from the LLS estimates may show high variance. A weight matrix estimated with low precision might lower the accuracy of the WLLS_2 estimator. Increasing the number of diffusion-weighted samples thus not only improves the precision of the WLLS_2 estimator, it also improves its accuracy. The latter does not hold for the WLLS_1. Recently, it has been suggested to use non-linear tensor estimators rather than its linear alternatives [Jones et al., 2013]. Our results, however, indicate the non-triviality in choosing between multi-step WLLS and NLS strategies for estimating dMRI parameters. Indeed, the decision depends on the number of diffusion-weighted signals (see Fig. 5.4), the parameter of interest [see Fig. 5.7 and Landman et al., 2007], and – possibly – the actual data distribution. However, further study is needed to extensively showcase the strengths, limitations, and pitfalls of the different diffusion estimators in case of non-Rice data distributions.

5.8 Summary

The DTI and DKI model have in common that they can be structured into a linear regression form depending on the natural logarithm of the diffusion-weighted

MR signals. The unknown model parameters can be estimated with an LLS estimator, or its weighted variants. Those estimators are widely used in DTI and DKI studies, because they come with several strengths. First, the (weighted) LLS estimators have a closed-form solution. Therefore, unlike iterative nonlinear strategies, the linear estimators are computationally efficient and not prone to getting stuck in a local optimum. Second, the linear estimators are potentially very accurate, especially compared to the NLS estimator [Salvador et al., 2005]. If the diffusion-weighted data is Rice distributed with SNR$>$ 2, then the linear estimators are even unbiased due to the zero expectation of the error term in the log-Rician framework. Third, the high accuracy is not at the expense of the ease of use. In other words, unlike many advanced diffusion parameter estimators, e.g., the maximum likelihood (ML) estimator, the linear estimators don't require the knowledge of the noise parameter σ_c [Veraart et al., 2012]. The estimation of the noise parameter is not only challenging, it also reduces the precision of the diffusion parameter estimator. Unfortunately, those advantages go hand in hand with some limitations (i.e. reduction of precision, optimal weights cannot be determined) and a potential pitfall (i.e. accuracy might strongly depend on selected weights).

Bibliography

S. Aja-Fernández and A. Tristán-Vega. Influence of noise correlation in multiple-coil statistical models with sum of squares reconstruction. *Magn Reson Med*, 67(2): 580–585, 2012.

S. Aja-Fernández, A. Tristán-Vega, and W.S. Hoge. Statistical noise analysis in GRAPPA using a parametrized noncentral Chi approximation model. *Magn Reson Med*, 65(4):1195–1206, 2011.

P.J. Basser, J. Mattiello, and D. Le Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B*, 103(3):247–254, 1994a.

P.J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophys J*, 66(1):259–267, 1994b.

P.A. Cook, Y. Bai, S. Nedjati-Gilani, K.K. Seunarine, M.G. Hall, G.J. Parker, and D.C. Alexander. Camino: Open-source diffusion-MRI reconstruction and processing. In *Proceedings 14th Scientific Meeting of the International Society for Magnc Reson Med*, page 2759, Seattle, USA, 2006.

M.A. Griswold, P.M. Jakob, R.M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*, 47(6):1202–1210, 2002.

M. Jenkinson, C.F. Beckmann, T.E.J. Behrens, M.W. Woolrich, and S.M. Smith. FSL. *NeuroImage*, 62(2):782 – 790, 2012.

J.H. Jensen, J.A. Helpern, A. Ramani, H. Lu, and K. Kaczynski. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magn reson imaging. *Magn Reson Med*, 53(6):1432–40, June 2005.

D.K. Jones and P.J. Basser. "squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. *Magn Reson Med*, 52(5):979–993, November 2004.

D.K. Jones, M.A. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magn reson imaging. *Magn Reson Med*, 42 (3):515–525, 1999.

D.K. Jones, T.R. Knösche, and R. Turner. White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. *NeuroImage*, 73:239–254, 2013.

C.G. Koay, L.C. Chang, J.D. Carew, C. Pierpaoli, and P.J. Basser. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *J Magn Reson*, 182(1):115–125, 2006.

B.A. Landman, S. Mori, and J.L. Prince. Systematic evaluation of linear and nonlinear DTI Estimation methods: an open framework. In *Proceedings 15th Scientific Meeting, International Society for Magnc Reson Med*, page 1522, Berlin, Germany, 2007.

J. Lätt, M. Nilsson, R. Wirestam, F. Stahlberg, N. Karlsson, M. Johansson, P. C. Sundgren, and D. van Westen. Regional values of diffusional kurtosis estimates in the healthy brain. *J Magn Reson Imaging*, 37(3):610–8, March 2013. ISSN 1522-2586.

D. Le Bihan and H. Johansen-Berg. Diffusion MRI at 25: Exploring brain tissue structure and function. *NeuroImage*, 61(2):324–341, 2012.

A Leemans and D K Jones. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn Reson Med*, 61(6):1336–1349, 2009.

A. Leemans, J. Sijbers, S. De Backer, E. Vandervliet, and P. Parizel. Multiscale white matter fiber tract coregistration: a new feature-based approach to align diffusion tensor data. *Magn Reson Med*, 55(6):1414–23, 2006.

A Leemans, B Jeurissen, J Sijbers, and Derek K Jones. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine*, page 3537, Honolulu, USA, 2009.

H. Lu, J.H. Jensen, A. Ramani, and J.A. Helpern. Three-dimensional characterization of non-Gsussian water diffusion in humans using diffusion kurtosis imaging. *NMR Biomed*, 19(2):236–247, 2006.

J. Mattiello, P.J. Basser, and D. Le Bihan. The b-matrix in diffusion tensor echo-planar imaging. *Magn Reson Med*, 37(2):292–300, 1997.

D.C. Noll, D.G. Nishimura, and A. Macovski. Homodyne detection in magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 10(2):154 –163, 1991.

S. Pieper, B. Lorensen, W. Schroeder, and R. Kikinis. The NA-MIC Kit: ITK, VTK, Pipelines, Grids and 3D Slicer as an Open Platform for the Medical Image Computing Community. In *Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 698–701, Arlington, USA, 2006.

C. Pierpaoli and P. J. Basser. Toward a quantitative assessment of diffusion anisotropy. *Magn Reson Med*, 36(6):893–906, 1996.

C. Pierpaoli, L. Walker, M. Irfanoglu, A. Barnett, P.J. Basser, L.C. Chang, C.G. Koay, S. Pajevic, G. Rohde, J. Sarlls, and M. Wu. TORTOISE: an integrated software package for processing of diffusion MRI data. In *Proceedings 18th Scientific Meeting, International Society for Magnetic Resonance in Medicine*, page 1597, Stockholm, Sweden, 2010.

G.K. Rohde, A.S. Barnett, P.J. Basser, S. Marenco, and C. Pierpaoli. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. *Magn Reson Med*, 51(1):103–114, 2004.

R. Salvador, A. Pena, D.K. Menon, T. A. Carpenter, J.D. Pickard, and E.T. Bullmore. Formal characterization and extension of the linearized diffusion tensor model. *Human Brain Mapping*, 24(2):144–155, 2005.

J. Sijbers, A. J. den Dekker, P. Scheunders, and D. Van Dyck. Maximum likelihood estimation of Rician distribution parameters. *IEEE Trans Med Imaging*, 17(3): 357–361, 1998.

A. Tristán-Vega, S. Aja-Fernández, and C.-F. Westin. Least squares for diffusion tensor estimation revisited: propagation of uncertainty with Rician and non-Rician signals. *NeuroImage*, 2011.

J. Veraart, D.H.J. Poot, W. Van Hecke, I. Blockx, A. Van der Linden, M. Verhoye, and J. Sijbers. More accurate estimation of diffusion tensor parameters using diffusion Kurtosis imaging. *Magn Reson Med*, 65(1):138–45, 2011. ISSN 1522-2594.

J. Veraart, J. Rajan, RR Peeters, A. Leemans, S. Sunaert, and J. Sijbers. A comprehensive framework for accurate diffusion MRI parameter estimation. *Magn Reson Med*, page doi: 10.1002/mrm.24529, 2012.

CHAPTER 6

Conditional least squares:

A practical alternative to Maximum Likelihood

Contents

The work in this chapter has been published as:
J. Veraart, J. Rajan, R. Peeters, A. Leemans, S. Sunaert, and J. Sijbers. Comprehensive framework for accurate diffusion MRI parameter estimation. *Magnetic Resonance in Medicine*, 2013, in press.

6.1 Introduction

Since many diffusion models require many diffusion-weighted MR images, often in combination with high b-values, dMRI is a technique that suffers from low SNR, severe geometric eddy current distortions, and long scanning times that cause subject motion to become more probable. The geometrical distortions and/or misalignments between diffusion-weighted images are generally minimized by a series of image processing steps preceding the diffusion model fitting. Furthermore, the final accuracy of the diffusion measures will largely depend on the appropriateness of the statistical model of the magnitude dMRI data during estimation of the diffusion measures. It is widely recognized that native magnitude MR data follows a Rice distribution under some well-defined conditions [e.g. Gudbjartsson and Patz, 1995, Dietrich et al., 2008, or see Discussion]. Hence, the assumption of the noise being additive and normally distributed as implicit in the use of ordinary least squares estimators cannot be made at low SNR [Gudbjartsson and Patz, 1995]. As a result, least squares estimators are most often biased in dMRI [e.g. Sijbers and den Dekker, 2004, Jones and Basser, 2004, Veraart et al., 2011]. Nevertheless, they are commonly preferred in dMRI analyses, probably because of their low computational expense. Furthermore, Salvador et al. [2005] showed that the log-transformed MR data (SNR > 2) can be closely approximated by a normal distribution with a variance equal to the reciprocal of the squares SNR. This property underlies the popularity of the weighted linear least squares estimator (WLLS) in combination with diffusion models that can be linearized by the log-transformation. Typical examples of such diffusion models are diffusion tensor imaging (DTI; [Basser et al., 1994]) and diffusion kurtosis imaging (DKI; [Jensen et al., 2005]). During the last decade, methods to further remove or reduce the noise bias in dMRI analyses have been presented. Jones and Basser [2004] modified the ordinary nonlinear least squares estimator (NLS; [Koay et al., 2006]) by replacing the diffusion model prediction by its approximate expectation under the assumption of a Rice noise model [Gudbjartsson and Patz, 1995]. Kristoffersen evaluated fitting the mean value of the Rice distribution to dMRI data to correct for the noise bias given repeated measurements in context of one-dimensional diffusion models [Kristoffersen, 2007]. Next, maximum likelihood (ML) estimation of diffusion model parameters was presented [e.g. Landman et al., 2007, Veraart et al., 2011]. Andersson [2008] further extended the ML approach to a maximum a posteriori (MAP) model by the introduction of priors on the model parameters. Both ML and MAP estimators depend on the knowledge of the full shape of the data distribution, whereas the methods presented in [Jones and Basser, 2004] and [Kristoffersen, 2007] only require the (approximate) expression of the expectation value of the data's distribution. The increased accuracy of those advanced diffusion estimators is typically observed in well-defined simulations, where, for instance, the assumptions regarding the data distribution are known to be valid. In practice, however, correcting for EPI distortions, subject motion, or eddy current induced geometric deformations alters the data distribution such that it can no longer be analytically expressed, potentially nullifying the consistency of the ML estimators. Therefore, we here describe and evaluate the conditional least squares estimator (CLS) in the context of (unaveraged) dMRI. Basically, the CLS is a nonlinear least squares estimator,

which recognizes a) the actual difference between the model prediction and its expectation value and b) the heteroscedasticity of the data due to the signal dependency of the variance. Therefore, the method is an asymptotically normal, consistent, and computationally convenient parameter estimator without the need for full description of the data distribution. We will evaluate the estimator in terms of accuracy, precision, and robustness to dMRI data correction applied prior to model fitting. Although the proposed estimator can be combined with all existing diffusion models, we limit ourselves in this study to the DTI and DKI models.

6.2 Maximum likelihood estimator

Consider N noisy magnitude diffusion-weighted signals $\tilde{S}_n$ $(n = 1, \cdots, N)$. If the data are independently noncentral χ distributed, then the probability of observing $\tilde{S}_n$, given its respective noise-free signal S_{0_n} and noise level σ_c is

$$p(\tilde{S}_n|S_{0_n}, \sigma_c) = \frac{\tilde{S}_n^L}{\sigma_c^2} S_{0_n}^{1-L} \exp\left(-\frac{S_{0_n}^2 + \tilde{S}_n^2}{2\sigma_c^2}\right) I_{L-1}\left(\frac{S_{0_n}\tilde{S}_n}{\sigma_c^2}\right), \tag{6.1}$$

with $S_{0_n} = \exp(\boldsymbol{\beta}\boldsymbol{B}_{n*})$ and $\boldsymbol{B}_{n*}$ the n^{th} row of the b-matrix $\boldsymbol{B}$. The diffusion model parameters $\boldsymbol{\beta}$ can be estimated voxelwise with a consistent, asymptotically normal and asymptotically efficient maximum likelihood estimator (MLE; [e.g. Sijbers et al., 1998, Veraart et al., 2011]) in each voxel by substituting the observed values for the stochastic variables and maximizing over the parameters:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{n=1}^{N} \ln p(\tilde{S}_n|S(\boldsymbol{\beta}, \boldsymbol{B}_{n*}), \sigma_c). \tag{6.2}$$

If (an estimate of) σ_c is not known in advance, σ_c can be estimated as part of the fitting procedure.

6.3 Conditional least squares estimator

An alternative estimation procedure is based on the minimization of a weighted sum of squared deviations about conditional expectations:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{n=1}^{N} W_{nn} \left[\tilde{S}_n - \mathrm{E}(\tilde{S}_n|S(\boldsymbol{\beta}, \boldsymbol{B}_{n*}), \sigma_c)\right]^2, \tag{6.3}$$

with W_{nn} the n^{th} element of the diagonal weight matrix $\boldsymbol{W}$, which equals the reciprocal of the variance of the measurement. After substituting m_0 in Eq. (3.19) by the predicted signal, the weight will depend on the diffusion model parameters. The weights should therefore iteratively be updated during the parameter estimation. The diffusion model parameters $\boldsymbol{D}$ are estimated by solving Eq. (6.3) using Levenberg-Marquardt optimization. In theory, σ_c can be estimated as part of the fitting procedure. However, the CLS without prior noise knowledge will lack consistency in case of isotropic diffusion as no unique solution then exists.

6.4 Strengths, limitations, and pitfalls

6.4.1 Theoretical and practical considerations

Theoretically, the MLE is a precise and accurate alternative to least squares estimation, irrespective of SNR. This nonlinear estimator has optimal asymptotical properties regarding accuracy and precision, but it requires the analytical expression of the data PDF. However, there are two practical concerns regarding the use of MLE in the context of dMRI. First, the analytical expression of the PDF is based on the noise level. Nowadays, the estimation of the noise level is challenging due to the use of parallel imaging techniques (see Chapter 7). Second, the necessity of data correction (e.g. motion and eddy current corrections) prior to model fitting causes the MLE's dependency on the data PDF to become a weakness because the altered data PDF can no longer be expressed analytically. Indeed, head motion and eddy current distortions can be reduced by realigning all diffusion-weighted images, $M_n (n = 1, ..., N)$, to a distortion-free reference image, often an nondiffusion-weighted image [Rohde et al., 2004]. Common practice is to use an affine approach:

$$\hat{M}_n = \frac{\boldsymbol{A}_n(M_n)}{\lambda_n} \tag{6.4}$$

with $\boldsymbol{A}_n$ the affine transformation and $\lambda_n = |J(\boldsymbol{A}_n)|$, the Jacobian determinant of $\boldsymbol{A}_n$. The $\boldsymbol{B}$-matrix should be reoriented accordingly [Leemans and Jones, 2009]. As it is most likely that the transformed voxel grids of the diffusion-weighted images do not coincide with any of the discrete voxel positions in the reference image, the diffusion-weighted images need to be resampled to the reference image grid. So, Eq. (6.4) basically corresponds to calculating image intensities at transformed voxel coordinates as a weighted sum of the scaled intensities at surrounding voxels. The scaling as well as the resampling will, however, alter the data distribution. First, the noncentral χ distribution is scale invariant, i.e. if the random variable m follows a noncentral χ distribution with noise-free signal ν and noise parameter σ_c, then $\frac{m}{\lambda_n}$ is also noncentral χ distributed with parameters $\frac{\nu}{\lambda_n}$ and $\frac{\sigma_c}{\lambda_n}$, respectively. The expectation value of the scaled distribution is scaled accordingly. Second, the *Central Limit Theorem* states that the average of a large set of independent and identically distributed samples tends to follow a normal distribution. Thus, interpolation between noncentral χ distributed samples, might cause the noncentral χ PDF to change *towards* a Gaussian PDF. No closed form expression exists for the resulting distribution. However, the expectation value of the native distribution does not alter in high SNR or homogeneous regions. This finding favorizes the use of the CLS as a practical alternative to MLE in dMRI. The assumption of zero-centered normally distributed error terms in regressive models render MLE equivalent to the minimization of a sum of squares. By accounting for the actual difference between the model prediction and its expectation value and the heteroscedasticity of the data due to the signal dependency of the variance of a noncentral χ distribution, the CLS is the most precise unbiased least squares estimator. The CLS enjoys consistency and asymptotic normality, under some mild regularity conditions [Newey and McFadden, 1994]. However, despite the data PDF will converge towards a normal distribution by motion or eddy current distortion corrections, the error terms are

generally not normally distributed if SNR is very low. Therefore, the CLS might have lower precision than the MLE.

6.4.2 Experimental validation

6.4.2.1 Single voxel simulations

Noncentral χ distributed data was simulated using the DTI model as well as the DKI model, given predefined diffusional tensor(s), b-values, and gradient directions. For the DTI model, MD and FA were $0.8 \times 10^{-3}\,\mathrm{mm^2/s}$ and 0.85, respectively. The b-value was $1000\,\mathrm{s/mm^2}$, and 60 gradient directions – isotropically distributed over a unit sphere using Coulomb's law of repulsion, were selected [Jones et al., 1999]. Since we also simulated 12 nondiffusion-weighted signals, a set of 72 intensities was obtained. For the DKI simulations, additional diffusion-weighted signals with $b = 2500\,\mathrm{s/mm^2}$ were simulated along the same gradient directions. The MK was 0.6223 for the single-voxel simulations. Noisy synthetic data was obtained by adding zero-mean Gaussian noise, $\mathcal{N}\left(0, \sigma_c^2\right)$, to both the real and imaginary part of the noise-free signals S_0. The composite magnitude signal $\tilde{S}$ was then obtained using the SoS method:

$$\tilde{S} = \sqrt{\sum_{l=1}^{L} \left|S_0 + \mathcal{N}_l\left(0, \sigma_c^2\right)\right) + j\mathcal{N}_l\left(0, \sigma_c^2\right)\right|^2}. \tag{6.5}$$

During all experiments, $L = 1$, unless explicitly mentioned otherwise.

First, Monte Carlo simulations (2500 trials) were done to evaluate the effect of varying $\mathrm{SNR}_{\mathrm{b_0}}$, that is, the SNR of the nondiffusion-weighted signal on the accuracy and precision of the (a) WLLS, (b) ordinary NLS, (c) MLE, and (d) CLS. The $\mathrm{SNR}_{\mathrm{b_0}}$ range was $[10\sqrt{L}, 20\sqrt{L}]$. The Gaussian noise level, σ_c, was assumed to be known a priori. In Fig. 6.1(a-c) the accuracy of the estimators is studied by plotting the FA, MD, and MD - derived from the average model parameters - as a function of the SNR. The bias inherent to ordinary NLS estimators is inversely proportional to the $\mathrm{SNR}_{\mathrm{b_0}}$. Properly accounting for the noise statistics, as done with MLE and CLS, minimized the bias. Although the accuracy of WLLS approximates the accuracy of the CLS and MLE in DTI analyses, the WLLS suffers from a lower precision Fig. 6.1(d-f). Note the slightly lower precision of CLS, compared to MLE, in the estimation of MK Fig. 6.1(f). The MSE in estimates of FA, MD, and MK were shown in Fig. 6.1(g-h) as function of the SNR.

Second, in Fig. 6.2, the effect of an incorrect noise prior – $\tilde{\sigma}_c = \sigma_c \times (1 +$ noise offset) – is shown in terms of accuracy and MSE in estimates of FA, MD, and MK. The SNR was 15. Noise offsets in a range of $[-100\% \cdots 100\%]$ were simulated. For $\tilde{\sigma}_c = 0$, the CLS is identical to the NLS. Therefore, the performance (both in terms of accuracy and MSE) of the CLS will converge to the performance of the NLS if σ_c is underestimated. The same findings were observed for the MLE. Obviously, the accuracy of the WLLS might exceed the accuracy of a badly initialized ($\tilde{\sigma}_c < \sigma_c$) CLS. Overestimation of σ_c, on the other hand, might be more harmful for the performance of the CLS. Starting from an overestimation of 25%, the MSE in estimates of FA and MD of the CLS might exceed the MSE of the WLLS and NLS.

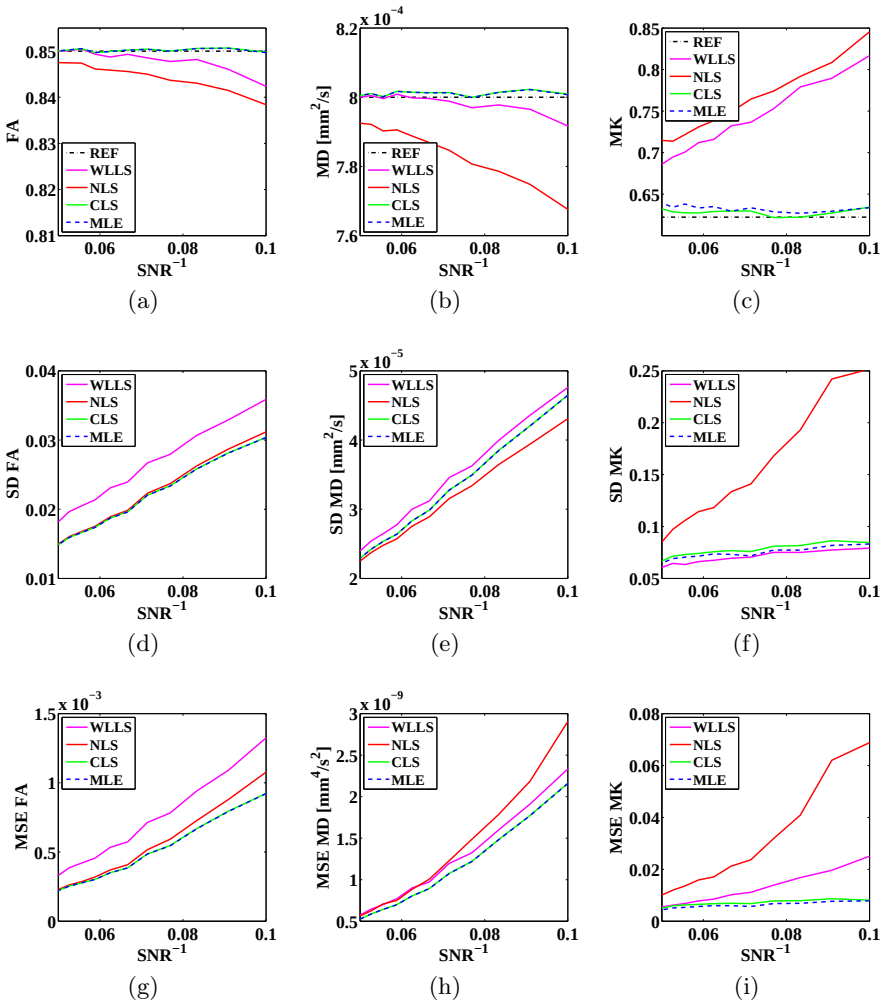
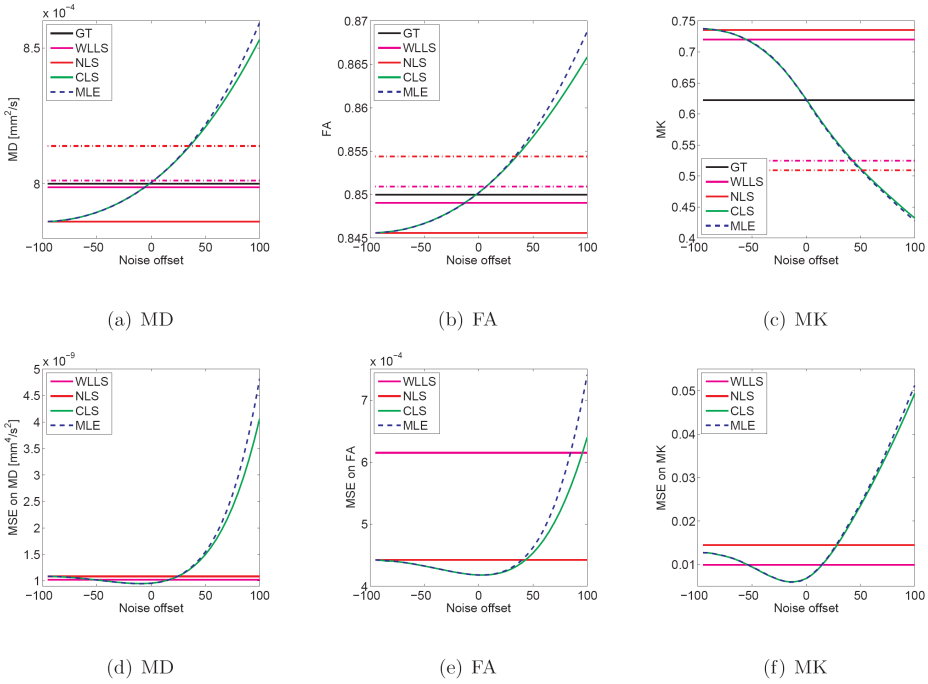


Figure 6.1: MD, FA and MK – FA, MD, and MK calculated from the average model parameters (a-c), the standard deviation (SD) of the estimated diffusion parameters (d-f), and MSE in the estimation of FA, MD, and MK (g-i) are shown as a function of the $\mathrm{SNR}_{b_0}^{-1}$ for WLLS, NLS, CLS and MLE.

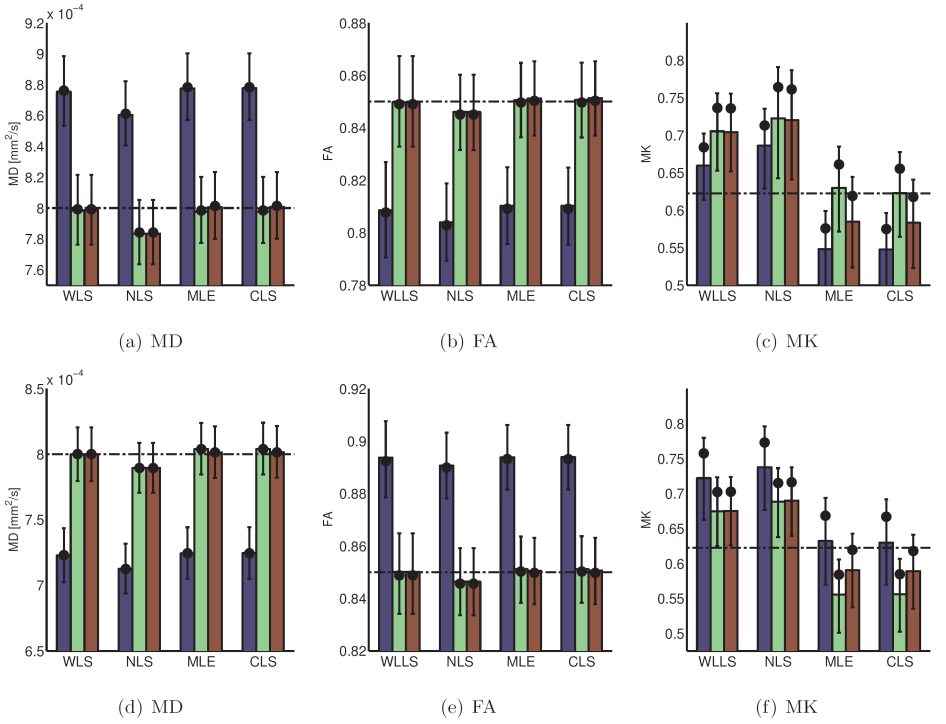


(a) MD (b) FA (c) MK

(d) MD (e) FA (f) MK

Figure 6.2: MD, FA and MK – calculated from the average diffusion(al) tensor(s) – and the MSE in estimates of MD, FA, and MK are shown as a function of an incorrect noise prior [%]. The red and pink dash-dotted lines in the upper graphs are the reflections of the red and pink solid lines across the black reference line.

The CLS performs only better than WLLS w.r.t. the MSE in estimates of MK if the noise offset is within the range: $[-50\%, +15\%]$. For MLE, similar conclusions can be drawn, however, the performance drops slightly more rapidly. Third, we quantified the effect of (not) applying the signal modulation to compensate for voxel volume changes that occur when reversing the stretch or compressing induced by the eddy currents on diffusion parameter estimation using WLLS, NLS, MLE, and CLS. Keeping in mind the scale invariance of the noncentral χ distribution, we also evaluated the need for accounting for the heteroscedasticity in the data, which is inherently induced by the signal modulation. Following simulation settings were chosen: 2500 trials, $L = 1$, $\mathrm{SNR}_{b_0} = 15$, maximal stretch for voxels was $\pm 8\%$ and $\pm 15\%$ for $b = 1000\,\mathrm{s/mm^2}$ and $b = 2500\,\mathrm{s/mm^2}$, respectively. In Fig. 6.3, it can be observed that not modulating the diffusion-weighted signal after eddy current correction will cause a significant bias in all estimators (blue). Intuitively, one might expect the bias to reduce when applying the signal modulation (green). However, our simulation results show that the bias can even increase (see Fig. 6.3(c)) when using NLS or WLLS. This is because the benefit of signal modulation can be opposed by an inflated noise bias (top figures for which $\lambda < 1$). The difference between the top and bottom ($\lambda > 1$) figures indicates that the quantitative effect of signal modulation depends largely on the sign of $\lambda - 1$. When using estimators



(a) MD (b) FA (c) MK

(d) MD (e) FA (f) MK

Figure 6.3: The bar plots show the effect of signal modulation (top: $\lambda < 1$, bottom: $\lambda > 1$) on MD, FA, and MK: no signal modulation (blue), signal modulation (green), signal + σ_c modulation (red). Height of the bars indicate the average diffusion parameter values, while the black dots are the parameter values derived from the averaged diffusion(al) tensor(s). The error bars span the interquartile range. The dashed horizontal black lines are the ground truth values.

that account for the noise statistics, it is clearly beneficial to scale the noise level accordingly (red). Taking the data's heteroscedasticity into account does not change the results for the WLLS and NLS significantly.

6.4.2.2 Whole brain simulations

The same simulated diffusion-weighted data sets were used to evaluate several estimation approaches: (a) WLLS, (b) NLS, (c) CLS with noise prior, (d) MLE with noise prior, (e) CLS without noise prior (CLS*), and (f) MLE without noise prior (MLE*). For (c) and (d), the noise prior was obtained using our proposed noise map estimator. For (e) and (f), the noise level was assumed to be unknown and estimated simultaneously with the diffusion model parameters. We evaluated the accuracy of all approaches in terms of commonly used diffusion parameters, i.e. FA, MD, and MK. The experiments were repeated after performing a half-pixel shift in both in-plane directions to introduce an interpolation effect, similar to the one that occurs when performing motion/eddy current distortion prior to model fitting. Both a DTI and DKI study was done. For the DTI study, the $b = 2500\,\mathrm{s/mm^2}$ shell

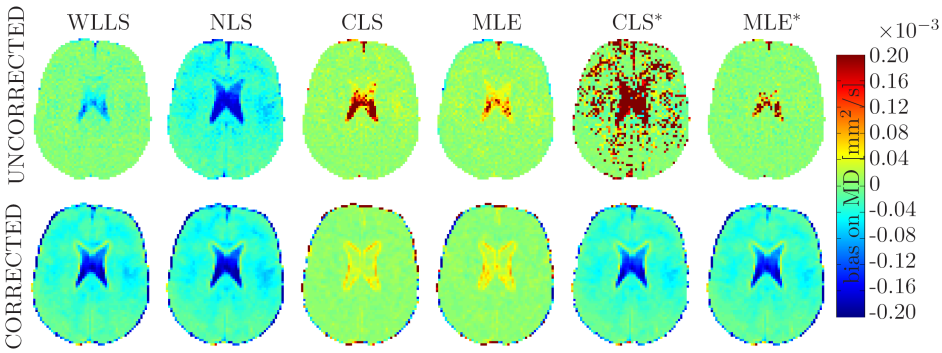


Figure 6.4: The DTI-estimators' biases on MD ($[mm^2/s]$) were shown. Following estimators were compared: WLLS, NLS, CLS, MLE, CLS*, and MLE*. The * indicates that the noise map estimation was a part of the fitting procedure. Results were shown without and with data resampling before model fitting in the top and bottom row, respectively

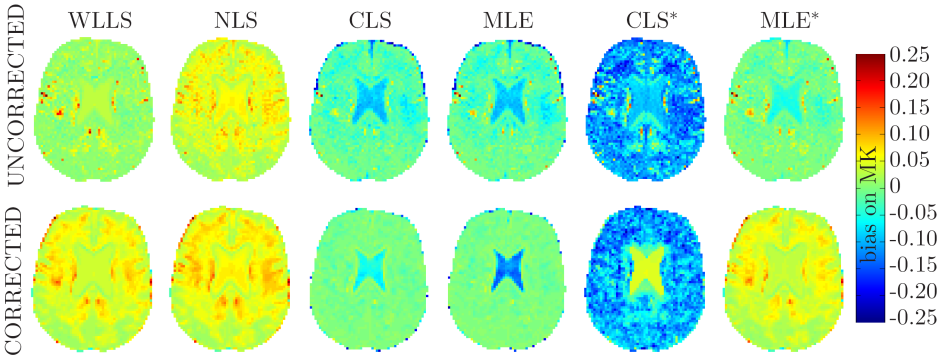


Figure 6.5: The DKI-estimators' biases on MK were shown. Following estimators were compared: WLLS, NLS, CLS, MLE, CLS*, and MLE*. Results were shown without and with data resampling before model fitting in the top and bottom row, respectively

was excluded from the fitting procedure. Fig. 6.6 shows the bias on MD for all DTI estimators, whereas Fig. 6.7 shows the bias on MK for the different DKI estimators. Without data resampling, the MLE, MLE* and CLS with noise prior are nearly unbiased, whereas the CLS* and NLS are clearly biased (top row). The bias of MLE, MLE* and CLS are caused by of the limited number of diffusion-weighted images in the study. Hence, the estimators cannot fulfill their asymptotic properties. Data interpolation, however, clearly changes the performance of the MLE* (bottom row). The central limit theorem dictates that the MLE without prior noise knowledge tends to behave like the biased NLS. The MLE with given noise level appears only minimally biased despite its less favorable properties. After data resampling, one can clearly notice the highly intense bias-rim surrounding the cerebrospinal fluid (CSF). Fortunately, similar *errors* are not (or much less) observed in other brain regions. It's worth noting the WLLS becomes less accurate in the estimation of MD and MK after data resampling.

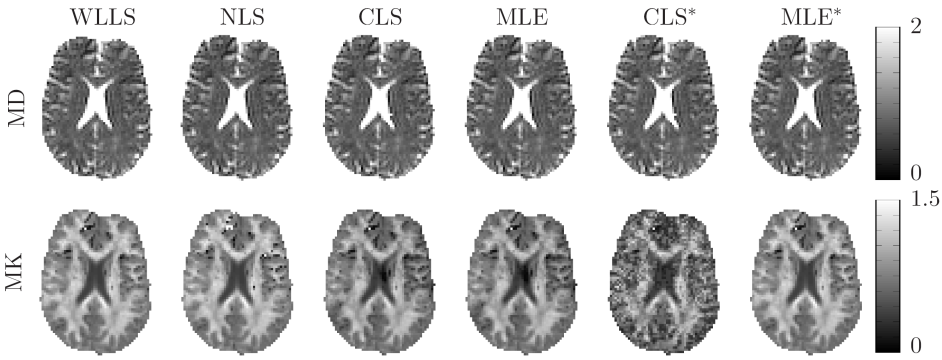


Figure 6.6: MD (top row; $[\times 10^{-3}\,\mathrm{mm}^2/\mathrm{s}]$) and MK (bottom row) maps obtained from the real data with the DTI and DKI model, respectively, are shown. The maps were obtained with different estimators: WLLS, NLS, CLS, MLE, CLS*, and MLE*. See Table 1 for quantitative differences.

6.4.2.3 Real data experimens

A 27 year-old healthy volunteer underwent imaging on a Siemens Trio (3T) MR scanner (Siemens AG, Siemens Medical Solutions, Erlangen, Germany), using a 12-channel receiver head coil. The body coil was used for transmission. A spin echo (SE) DWI sequence was used in acquiring the dMRI data. Diffusion weighting was applied along 60 isotropically distributed gradient directions with $b = 1000\,\mathrm{s/mm}^2$ as well as $b = 2500\,\mathrm{s/mm}^2$. Additionally, 12 images without diffusion sensitization were acquired. Other imaging parameters were kept constant throughout the DKI data acquisition sequences: TR/TE : 6100/118 ms, matrix: 96×96, NEX: 1, slice thickness: 2.5 mm, slices: 40, parallel imaging factor: mSENSE with acceleration factor 2. diffusion-weighted data was corrected for motion and eddy current distortions using exploreDTI [Leemans et al., 2009]. A DTI and DKI analyses were done. For DTI in particular, the $b = 2500\,\mathrm{s/mm}^2$ shell was additionally excluded. Preceding the DKI fit, the Gibbs phenomenon was reduced by isotropic smoothing of the diffusion-weighted data (FWHM = 3 mm). The quantitative differences between the different estimators w.r.t. MD, FA, and MK were evaluated with an ROI analysis. Fig. 6.6 shows the MD maps (*top row*) and MK maps (*bottom row*), estimated with the DTI and DKI model, respectively, using the different estimators. After manual delineation of the corpus callosum in a single slice, an ROI analysis of the diffusion measures was performed (see Table 6.1). The paired Wilcoxon signed rank test was applied to evaluate the pairwise differences between the estimated diffusion measures. Besides a single exception, all differences were statistical significant ($p < 0.05$). Only no statistical difference in terms of MD was observed when comparing CLS and MLE ($p = 0.39$).

Table 6.1: Real data ROI analysis of the diffusion parameters (FA, MD [$\times 10^{-5}$ s/mm^2], and MK), based on a manual delineation of the corpus callosum.

	WLLS	NLS	CLS	MLE	CLS*	MLE*
FA	0.68	0.68	0.69	0.69	0.70	0.68
MD	79.73	78.99	81.16	81.12	81.78	80.99
MK	1.10	1.10	1.00	0.99	0.95	1.09

6.5 Discussion

CLS v. MLE

The MLE has some desirable properties such as consistency, and asymptotic normality. The desirable statistical properties of MLE stems from its basis on the joint PDF of the data. Simulations showed that the same asymptotic properties can be attained by the CLS, an estimator that does not require full knowledge of the data's joint PDF. The CLS only requires an analytical expression of the first moments of the data distribution. Compared to MLE, CLS is easier to implement and, more importantly, computationally far less intensive. From our current MATLAB implementations, we observed that CLS is about 50× faster than MLE. With CLS, the whole brain DTI analysis performed as part of the real data experiment took less than 3 minutes on a 64-bit quad-core computer, each core running at 2.80 GHz. An additional argument to prefer CLS to MLE roots in the necessity of data correction prior to model fitting. The data correction (e.g. motion and eddy current corrections) causes the MLE's dependency on the joint PDF to become a weakness because the altered data PDF can no longer be expressed in closed form. So, theoretically, the MLE loses then its consistency, whereas the CLS might keep that property as the expectation value of the distribution will not alter during data processing in high SNR or homogeneous regions. In practice, the MLE's drop in accuracy is SNR dependent, but - for moderate SNR - it has been shown that it is fairly limited if the native noise parameter is known in advance (see Fig. 6.4 and Fig. 6.5).

CLS v. NLS

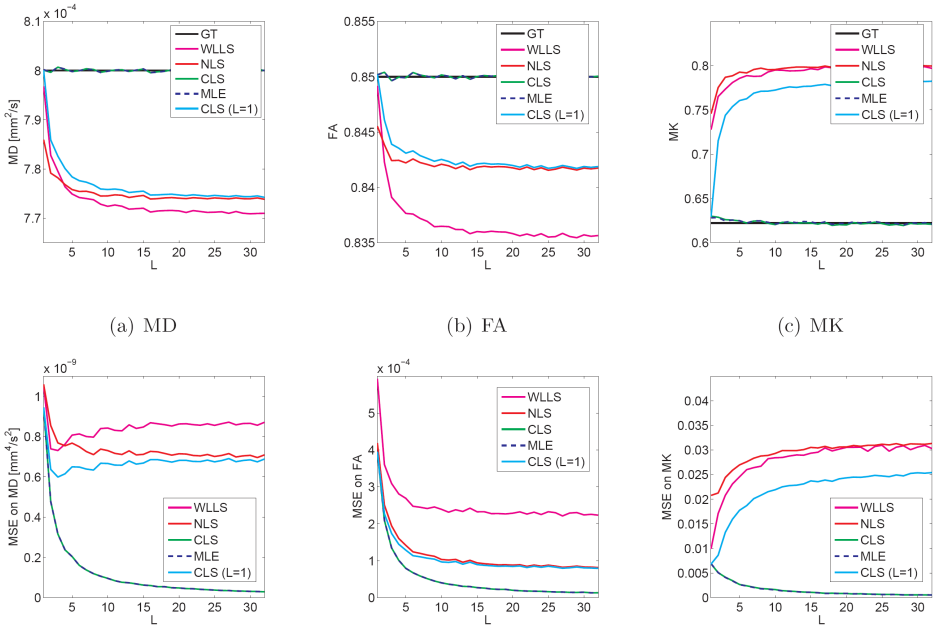
Basically, the CLS is a nonlinear least squares estimator, which recognizes a) the actual difference between the model prediction and its expectation value and b) the heteroscedasticity of the data due to the signal dependency of the variance of a noncentral χ distribution. In the ordinary NLS (as described in [Koay et al., 2006]), both (a) and (b) are neglected because of the assumption that a noncentral χ distribution can be well approximated by a normal distribution centered around the noise-free signal. That assumption, however, is violated at low SNR as a result of which NLS lacks accuracy in dMRI analyses. An SNR dependent, systematic error in the calculation of clinically relevant diffusion parameters is observed with the NLS, while the CLS is asymptotically unbiased. At high SNR, both estimators are identical. Difference in computation time between both strategies is minimal if a lookup table is used for the evaluation of the hypergeometric function in Eq. (3.17).

CLS v. WLLS

Some diffusion models, such as DTI and DKI, can easily be linearized by a log-transformation, as a result of which fast linear LS estimators has become popular for parameter estimation. Since Salvador et al. [2005] showed that the log-transformed Rice distributed MR data (SNR > 2) can be closely approximated by a zero-centered distribution with a variance equal to the reciprocal of the squared SNR, the WLLS is expected to be more accurate than the NLS. For clinically relevant SNR values, the WLLS was shown to be nearly as accurate as the CLS and MLE in the estimation of DTI model parameters. However, bear in mind following remarks: (a) the improved accuracy, compared to ordinary NLS, vanishes if magnitude operations are applied prior to model fitting. Because of the changing data distribution, the mathematical reasoning of Salvador et al. [2005] no longer holds; (b) Overall, the MSE of the WLLS is high compared to CLS. Obviously, the WLLS outperforms all other methods in terms of computational cost.

Data distribution

Aja-Fernandez and colleagues showed that in case of parallel imaging, the magnitude MR data follows (approximately) a noncentral χ distribution with $2L$ degrees of freedom, in which L is the effective number of receiver coils [Aja-Fernández et al., 2011, Aja-Fernández and Tristán-Vega, 2012]. The Rice distribution is only a special case ($L = 1$). The knowledge of L is essential, though often nontrivial because of its dependency to the applied imaging technique, the actual number of receiver channels and reconstruction filter. It is of utmost importance to grasp that L can only readily be determined in following scenarios: (a) nonparallel imaging with negligible coil correlations; L equals the actual number of receiver coils [Constantinides et al., 1997], (b) parallel imaging with image-domain reconstruction; L equals 1 [Dietrich et al., 2008], and (c) parallel imaging with a reconstruction filter such as spatially matched filters or an adaptive reconstruction [Dietrich et al., 2008]. These filters will create a single complex image by linearly combining the complex signals from all receiver channels before calculating the magnitude. Again, L equals 1. Frequency domain reconstruction (e.g. GRAPPA; [Griswold et al., 2002]) and nonparallel imaging for which the coil correlations cannot be neglected, are far more difficult to model after SoS reconstruction of the composite magnitude data [Aja-Fernández et al., 2011, Aja-Fernández and Tristán-Vega, 2012]. In particular, the complex data and internal reconstruction parameters are required for the estimation of L. Therefore, the actual data distribution cannot be determined from magnitude MR data only. The use of more advanced diffusion estimators thus remains challenging in those cases, whereas their importance increases. The bias of the ordinary NLS and WLLS is proportional to L (see Fig. 6.7). Note that we preferred to express the MLE and CLS as a function of L for the purpose of generalization. Basically, if L is known, the MLE and CLS can be used regardless the value of L or the applied imaging technique.



(a) MD (b) FA (c) MK

(d) MD (e) FA (f) MK

Figure 6.7: MD, FA and MK – calculated from the average diffusion(al) tensor(s) – and the MSE in estimates of MD, FA, and MK are shown as a function of the number of receiver channels, L. During simulations (cfr. section 6.4.2.1) L varied from 1 to 32, while σ_c was kept constant ($\text{SNR}_{b_0} = 15$). The performance of estimators, which not properly account for the actual data distribution (NLS, WLLS and estimators designed for Rice distribution, such as CLS with $L = 1$) decreases with L.

Physiological noise

The quest for accuracy will unceasingly be hindered by the presence of physiological noise. Spatially and temporally varying artifacts, e.g. cardiac pulsation or system instabilities, will perturb the diffusion-weighted signals in such a way they cannot be modeled. More robust estimators, i.e. estimators that are less sensitive to the presence of outliers, are needed to deal with those perturbations. A more robust implementation of the CLS using the (informed) RESTORE approach is trivial [Chang et al., 2005, 2012].

6.6 Summary

During the last decade, many approaches have been proposed for improving the estimation of diffusion measures. These techniques have already shown an increase in accuracy based on theoretical considerations, such as incorporating prior knowledge of the data distribution. The increased accuracy of diffusion metric estimators is typically observed in well-defined simulations, where the assumptions regarding properties of the data distribution are known to be valid. In practice, however, correcting for subject motion and geometric eddy current deformations alters the data distribution tremendously such that it can no longer be expressed in a closed form. The image processing steps that precede the model fitting will render several assumptions on the data distribution invalid, potentially nullifying the benefit of applying more advanced diffusion estimators. In this work, we present a generic diffusion model fitting framework that considers some statistics of diffusion MRI data. In this chapter, we introduced the CLS as a practical alternative to the MLE. We demonstrated that the accuracy of that particular estimator can generally be preserved, regardless the applied preprocessing steps, if the noise parameter is known a priori. To fulfill that condition, we will propose an approach for the estimation of spatially varying noise levels in Chapter 7.

Bibliography

S. Aja-Fernández and A. Tristán-Vega. Influence of noise correlation in multiple-coil statistical models with sum of squares reconstruction. *Magn Reson Med*, 67(2): 580–585, 2012.

S. Aja-Fernández, A. Tristán-Vega, and W.S. Hoge. Statistical noise analysis in GRAPPA using a parametrized noncentral Chi approximation model. *Magn Reson Med*, 65(4):1195–1206, 2011.

J. Andersson. Maximum a posteriori estimation of diffusion tensor parameters using a Rician noise model: Why, how and but? *NeuroImage*, 42(4):1340–1356, 2008.

P.J. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophys J*, 66(1):259–267, 1994.

L.-C. Chang, D.K. Jones, and C. Pierpaoli. RESTORE: Robust estimation of tensors by outlier rejection. *Magn Reson Med*, 53(5):1088–1095, 2005.

L.-C. Chang, L. Walker, and C. Pierpaoli. Informed RESTORE: A method for robust estimation of diffusion tensor from low redundancy datasets in the presence of physiological noise artifacts. *Magn Reson Med*, 86(5):1654–1663, 2012.

C.D. Constantinides, E. Atalar, and E.R. McVeigh. Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magn Reson Med*, 38(5):852–857, 1997.

O. Dietrich, J.G. Raya, S.B. Reeder, M. Ingrisch, M.F. Reiser, and S.O. Schoenberg. Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics. *Magn Reson Imaging*, 26(6):754–762, 2008.

M.A. Griswold, P.M. Jakob, R.M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*, 47(6):1202–1210, 2002.

H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magn Reson Med*, 34(6):910–914, 1995.

J.H. Jensen, J.A. Helpern, A. Ramani, H. Lu, and K. Kaczynski. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magn reson imaging. *Magn Reson Med*, 53(6):1432–40, June 2005.

D.K. Jones and P.J. Basser. "squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. *Magn Reson Med*, 52(5):979–993, November 2004.

D.K. Jones, M.A. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magn reson imaging. *Magn Reson Med*, 42 (3):515–525, 1999.

C.G. Koay, L.C. Chang, J.D. Carew, C. Pierpaoli, and P.J. Basser. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *J Magn Reson*, 182(1):115–125, 2006.

A. Kristoffersen. Optimal estimation of the diffusion coefficient from non-averaged and averaged noisy magnitude data. *J Magn Reson Imaging*, 187(2):293–305, August 2007.

B. Landman, P.-L. Bazin, and J. Prince. Diffusion tensor estimation by maximizing Rician likelihood. In *IEEE 11th International Conference on Computer Vision, 2007.*, pages 1 –8, Rio de Janeiro, Brazil, oct. 2007.

A Leemans and D K Jones. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn Reson Med*, 61(6):1336–1349, 2009.

A Leemans, B Jeurissen, J Sijbers, and Derek K Jones. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine*, page 3537, Honolulu, USA, 2009.

W.K. Newey and D. McFadden. Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111 – 2245. Elsevier, 1994.

G.K. Rohde, A.S. Barnett, P.J. Basser, S. Marenco, and C. Pierpaoli. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. *Magn Reson Med*, 51(1):103–114, 2004.

R. Salvador, A. Pena, D.K. Menon, T. A. Carpenter, J.D. Pickard, and E.T. Bullmore. Formal characterization and extension of the linearized diffusion tensor model. *Human Brain Mapping*, 24(2):144–155, 2005.

J. Sijbers and A.J. den Dekker. Maximum likelihood estimation of signal amplitude and noise variance from mr data. *Magn Res Med*, 51(3):586–594, 2004.

J. Sijbers, A. J. den Dekker, P. Scheunders, and D. Van Dyck. Maximum likelihood estimation of Rician distribution parameters. *IEEE Trans Med Imaging*, 17(3): 357–361, 1998.

J. Veraart, W. Van Hecke, and J. Sijbers. Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a Rician noise model. *Magn Reson Med*, 66(3):678–686, 2011.

CHAPTER 7

Estimation of spatially variable Rician noise map

Contents

The work in this chapter has been published as:
J. Veraart, J. Rajan, R. Peeters, A. Leemans, S. Sunaert, and J. Sijbers. Comprehensive framework for accurate diffusion MRI parameter estimation. *Magnetic Resonance in Medicine*, 2013, in press.

7.1 Introduction

The likelihood function or the expectation value of a Rice distributed variable depend on the underlying Gaussian noise level, i.e. the standard deviation of the real and imaginary parts of the complex signal σ_c. Although some methods suggested estimating the Gaussian noise level as part of the model fit, the Gaussian noise level is preferably estimated prior to model fitting because the accuracy and precision of diffusion model parameter estimators will depend on the prior knowledge of this noise level. Most of the existing noise estimation methods can be classified as methods that use (a) background areas [e.g. Sijbers et al., 2007, Chang et al., 2005, Koay et al., 2009b] or (b) the image object itself [e.g. Coupé et al., 2010, Maximov et al., 2012, Landman et al., 2009, Manjón et al., 2010, Koay and Basser, 2006] to estimate the noise variance. For a thorough overview, see [Aja-Fernández et al., 2009]. The background-based methods often fail due to the suppression of the background signal by the scanner or spatially varying noise. The object-based methods often rely on (a) a Gaussian approximation of the noise [e.g. Landman et al., 2009], (b) a sufficiently high spatial resolution such that a (non)local set of voxels with similar neighborhoods can be found to fit a noise distribution [e.g. Manjón et al., 2010], (c) repeated measurements [e.g. Sijbers et al., 1998, Maximov et al., 2012, Landman et al., 2009, Koay and Basser, 2006], or (d) a spatially uniform distribution of the noise level [e.g. Coupé et al., 2010, Rajan et al., 2010]. dMRI, however, suffers from a restricted spatial resolution and involuntary subject and/or brain motion causing misalignments between multiple measurements. Furthermore, the noise is generally spatially varying due to the use of parallel imaging techniques [Robson et al., 2008]. We here propose a new strategy that allows for the voxelwise estimation of the noise level. The strategy is a dMRI-specific extension of the work of Coupé et al. [2010], who combined the Median Absolute Deviation (MAD) estimator in the wavelet domain with an iterative correction scheme, as proposed by Koay and Basser [2006].

7.2 Noise level estimation

Coupé et al. [2010] showed that for stationary noise, a robust estimate of the standard deviation of the magnitude MR signal, $\hat{\sigma}_{\text{M}}$, can be obtained:

$$\hat{\sigma}_{\text{M}} = \frac{\text{median}\left(|\boldsymbol{y}|\right)}{0.6745}, \tag{7.1}$$

with $\boldsymbol{y}$ the wavelet coefficients of the highest sub-band HHH, which is mainly composed of the coefficients corresponding to the noise [Donoho, 1995]. An estimate of σ_c can then be calculated as:

$$\hat{\sigma}_c = \frac{\hat{\sigma}_{\text{M}}}{\sqrt{\xi_\theta}}, \tag{7.2}$$

with

$$\xi_\theta = 2L + \theta^2 - \beta_L^2 \left[{}_1F_1\left(-\frac{1}{2}, L, -\frac{\theta^2}{2}\right)\right]^2, \tag{7.3}$$

a correction term that depends on the SNR, $\theta = \frac{\nu}{\sigma_c}$ [Koay and Basser, 2006]. Because the SNR is not known a priori, the correction term needs to be estimated iteratively until convergence, initiated by both the mean, μ_m, and variance, σ_m^2, of the magnitude signal. An extension to the estimation of a 3D noise map, $\sigma_c(\boldsymbol{x})$ with $\boldsymbol{x}$ the 3D spatial location, relies on the observation that the median and monotone functions, such as $\frac{1}{\sqrt{\xi(\theta)}}$ in Eq. (7.2), commute. More specifically, an initial estimate of the magnitude noise map $\hat{\sigma}_{m,n}(\boldsymbol{x})$ of the n^{th} diffusion-weighted volume is obtained as follows:

$$\hat{\sigma}_{\mathrm{M},n}(\boldsymbol{x}) = \frac{|y_n(\boldsymbol{x})|}{0.6745}. \tag{7.4}$$

In Eq. (7.4), $y_n(\boldsymbol{x})$ denotes the coefficient at location $\boldsymbol{x}$ of the HHH sub-band of an undecimated wavelet decomposition of the n^{th} uncorrected diffusion-weighted volume. Because $\hat{\sigma}_{\mathrm{M},n}(\boldsymbol{x})$ will vary with n – due to the varying underlying diffusion-weighted signals – one cannot instantly apply the median operator along n to obtain a robust estimate of $\hat{\sigma}_{\mathrm{M}}(\boldsymbol{x})$. Next, the magnitude signal average at location $\boldsymbol{x}$, $\hat{\mu}_{\mathrm{M},n}(\boldsymbol{x})$, can be estimated by fitting a spherical spline model to a single b-valued diffusion-weighted dataset composed of N_{B} diffusion-weighted images using linear regression [Koay et al., 2009a]. In case of severe head motion, the spherical spline model is preferably fitted to the motion corrected diffusion-weighted images. The reconstructed diffusion-weighted signals, $\hat{\mu}_{\mathrm{M},n}(\boldsymbol{x})$, should afterwards be resampled to their native spaces. Given $\hat{\theta}_n = \frac{\hat{\mu}_{\mathrm{M},n}(\boldsymbol{x})}{\hat{\sigma}_{\mathrm{M},n}(\boldsymbol{x})}$, $\xi_{\theta,n}(\boldsymbol{x})$ can be obtained using Eq. (7.3). Finally, the 3D noise map can be calculated:

$$\sigma_c(\boldsymbol{x}) = \mathrm{median}\left\{ \frac{\hat{\sigma}_{\mathrm{M},n}(\boldsymbol{x})}{\sqrt{\xi_{\theta,n}(\boldsymbol{x})}} | n = 1 \cdots N_{\mathrm{B}} \right\} \tag{7.5}$$

Note that $\sigma(\boldsymbol{x})$ is presumed to be spatially varying within each uncorrected diffusion-weighted volume, but, for a given $\boldsymbol{x}$, constant across the N_{B} uncorrected diffusion-weighted volumes.

7.3 Experiments

7.3.1 Simulation experiment

We simulated whole brain diffusion-weighted data, acquired with a 8-channel head coil and reconstructed with mSENSE (acceleration factor $R = 2$) using PULSAR [Jim et al., 2007]. Phantom images were derived from a hybrid diffusion atlas [Dhollander et al., 2011]. In addition to 12 non-diffusion-weighted volumes, 2 q-shells were sampled from the atlas: *Jones60* directions at $b = 1000\,\mathrm{s/mm^2}$ and $b = 2500\,\mathrm{s/mm^2}$. DKI model errors were removed by fitting the DKI model to the data and recalculating the diffusion-weighted volumes afterwards. Multichannel k-space data was slice-by-slice calculated as the Fourier transform of each individual coil image, which is the reconstructed diffusion-weighted image modulated by a normalized coil sensitivity map. Complex, Gaussian noise was added to the k-space data and all odd phase encoding k-space lines were suppressed before mSENSE

reconstruction [Wang et al., 2001]. We used these simulated data sets to evaluate the accuracy of our noise map estimator. The nonuniform reference noise map, σ_{SENSE}, was derived from the geometry factor map [Pruessmann et al., 1999]. The experiment was repeated for multiple SNR_{b_0} values. The SNR_{b_0} is here defined as the ratio between the median nondiffusion-weighted signal and the median $\sigma_{\text{SENSE}}(\boldsymbol{x})$. The proposed noise map estimation approach was applied on either of the b-value shells.

7.3.2 Real data experiment

A 27 year-old healthy volunteer underwent imaging on a Siemens Trio (3T) MR scanner (Siemens AG, Siemens Medical Solutions, Erlangen, Germany), using a 12-channel receiver head coil. The body coil was used for transmission. A spin echo (SE) DWI sequence was used in acquiring the dMRI data. Diffusion weighting was applied along 60 isotropically distributed gradient directions with $b = 1000\,\text{s/mm}^2$ as well as $b = 2500\,\text{s/mm}^2$. Additionally, 2×12 additional images were acquired: (a) 12 images without diffusion sensitization and (b) 12 images with diffusion sensitization ($b = 3000\,\text{s/mm}^2$) along a single gradient direction. Other imaging parameters were kept constant throughout the DKI data acquisition sequences: TR/TE : 6100/118 ms, matrix: 96×96, NEX: 1, slice thickness: 2.5 mm, slices: 40, parallel imaging factor: mSENSE with acceleration factor 2. diffusion-weighted data was corrected for motion and eddy current distortions using exploreDTI [Leemans et al., 2009].

Because of a lack of a ground truth, we had to define a bronze standard, i.e. a heuristic reference map, on the real data to evaluate our noise map estimation approach. As a bronze standard, we adopted the strategy proposed by Maximov et al. [2012]. Their strategy relies on the acquisition of repeated measurements and can, thus, be applied on the $b = 0\,\text{s/mm}^2$ as well as on the $b = 3000\,\text{s/mm}^2$ diffusion-weighted images. Our method was subsequently applied on the $b = 1000\,\text{s/mm}^2$ and $b = 2500\,\text{s/mm}^2$ shell.

7.4 Results

7.4.1 Simulation experiment

The estimated noise maps were averaged over 50 trials and shown in Fig 7.1. A slight positive bias can be observed. For the DTI set-up, i.e. using only the $b = 1000\,\text{s/mm}^2$ shell was used for the noise map estimation, the bias was 3%, 2%, 1.6%, and 1.5% for SNR_{b_0} =16, 12, 10, and 8, respectively. When using only the $b = 2500\,\text{s/mm}^2$ shell for noise map estimation, the bias was 1.6%, 1.6%, 2%, and 2.5% for SNR_{b_0}=16, 12, 10, and 8, respectively. The resulting noise maps are clearly much more accurate than the noise maps estimated as part of the CLS* and MLE* fitting procedures (see Fig. 7.2)

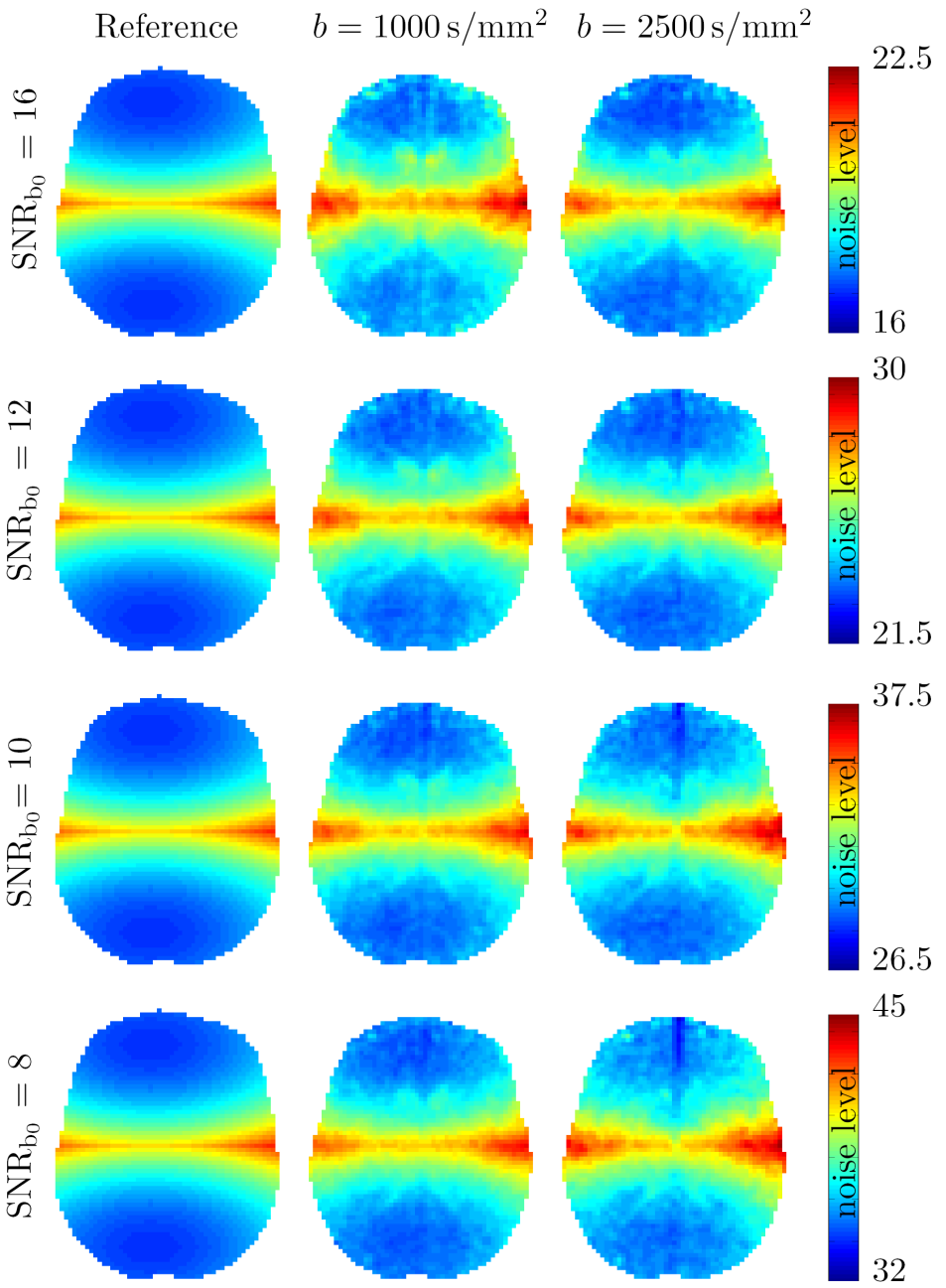


Figure 7.1: In the left column, the reference noise maps were shown. The average noise maps using the proposed estimator based on the diffusion-weighted images with $b = 1000\,\mathrm{s/mm^2}$ and $b = 2500\,\mathrm{s/mm^2}$ were shown in the middle and right column, respectively. Contrast was kept constant for visual comparison purposes.

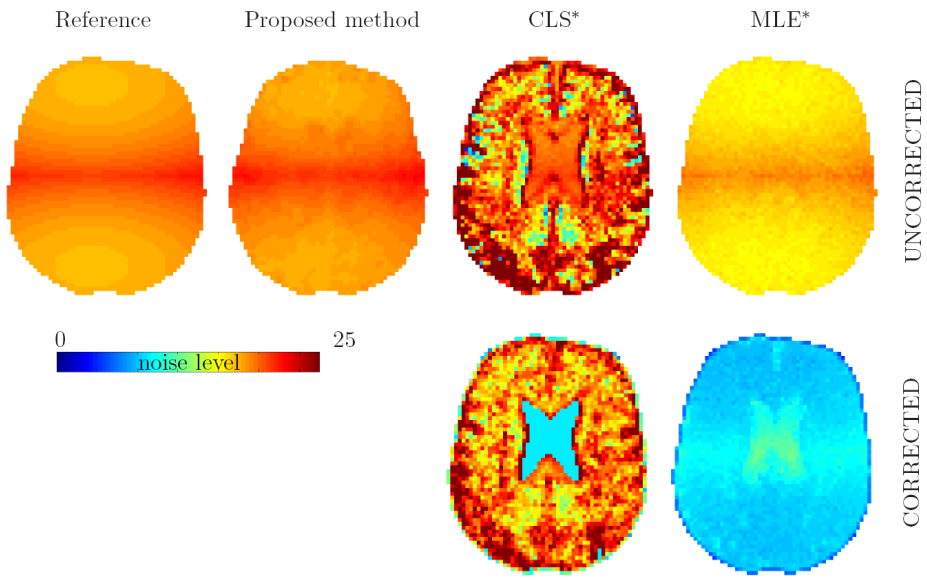


Figure 7.2: Noise map estimation as part of the CLS and MLE fitting procedures, with (bottom) and without (top) preceding data preprocessing, was compared to the reference noise map and the one obtained with our proposed method. The shown maps are average noise maps, calculated over 50 simulation trials.

7.4.2 Real data experiment

In a first experiment, the assumption of a constant noise level across the diffusion-weighted images was validated. For each diffusion-weighted image, the Gaussian noise level was calculated in 10 predefined background regions ($\nu = 0$), each including 300 voxels using Eq. (3.17). Trends in the resulting curves (i.e. σ as a function of the diffusion-weighted index) were detected with second-order polynomial regression. The first- and second-order terms were centered around 0, and thus, justified the assumption (results not shown). Next, Fig. 7.3 shows several noise maps. The maps are 2D noise maps, calculated as the average of ten consecutive slices within the estimated 3D noise maps. The averaging was mainly of importance to improve the precision of the bronze standard. Our proposed method resulted in noise maps (Fig. 7.3(a,b)) that correspond well with the bronze standard, both in terms of intensity and spatial distribution, especially if we compare our results with the bronze standard calculated from the $b = 3000\,\mathrm{s/mm^2}$ images (Fig. 7.3(d)). The noise map derived from the $b = 0\,\mathrm{s/mm^2}$ images (Fig. 7.3(c)) is clearly affected by pulsation artifacts [Tournier et al., 2011]. This is mainly reflected in an increased noise level in regions surrounding the CSF. Strong edges (i.e. high frequent image information) in the diffusion-weighted images will show-through in the HHH sub-band. The effect is less pronounced if low SNR images (e.g., with high b-values) are used for noise map estimation. Compare, for instance, Fig. 7.3(a) and Fig. 7.3(b), which are the noise maps estimated from the $b = 1000\,\mathrm{s/mm^2}$ and $b = 2500\,\mathrm{s/mm^2}$ images, respectively. Although suppression of the high gradient

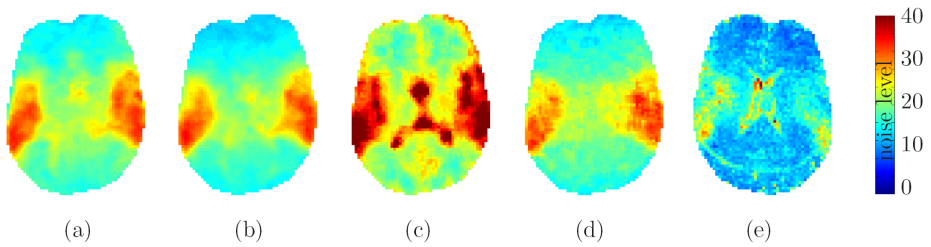


Figure 7.3: Our proposed noise map estimator, applied to the $b = 1000\,\mathrm{mm}^2/\mathrm{s}$ images (a) and $b = 2500\,\mathrm{mm}^2/\mathrm{s}$ images (b), was compared to a bronze standard, calculated from the $b = 0\,\mathrm{mm}^2/\mathrm{s}$ repetitions (c) or $b = 3000\,\mathrm{mm}^2/\mathrm{s}$ repetitions (d). The noise map obtained with MLE* (e) was shown for comparison.

regions is encouraged to minimize an artificial increase of noise level in those regions [Coupé et al., 2010], entirely removing the effect remains challenging and might require masking out of CSF regions. The noise map shown in Fig. 7.3(e) was obtained with the DTI-MLE*. The underestimated noise levels are in agreement with the simulations.

7.5 Discussion

Key for data modeling is the noise parameter. Since estimating the noise level as part of the fitting procedure appeared inaccurate, especially after applying data correction, the development of a 3D noise map estimation strategy is a main contribution of the work. For the real data experiments, the estimated noise map corresponds very well with the bronze standard, which was constructed with a recently proposed technique based on repeated measurements [Maximov et al., 2012]. In a clinical setting, that approach, however, has two clear limitations: (a) the acquisition of repeated measurements (preferably with high b-value) will further lengthen the scan time, (b) misalignment between repeated measurements, which will cause the bronze standard to become erroneous, is much more likely to occur for a patient than for an instructed healthy volunteer. For noise map estimation, the highest b-value shell is preferably used. In those images, the HHH sub-band is the least corrupted with residual signal, which originates in high image gradients. One might benefit from further suppressing the HHH sub-band's residual signal. Coupé et al. [2010] already suggested to threshold the gradient magnitude of LLL sub-band to create a mask for high gradient regions. Note that the proposed noise map estimator relies on the assumption of an identical Gaussian noise map for all uncorrected diffusion-weighted images. This assumption is violated by frequency domain reconstructions that are subsequently followed by SoS. The noncentral χ distribution will then be parameterized by an effective noise level, which depends on the underlying signal [Aja-Fernández et al., 2011, Aja-Fernández and Tristán-Vega, 2012]. However, the calculation of L provides the user already with the effective noise map. Obviously, if the underlying noise level is spatially invariant, then robust and precise background-based techniques might be preferred as the background

will comprise many voxels sampled from the same distribution [Sijbers et al., 2007, Chang et al., 2005, Koay et al., 2009b].

Bibliography

S. Aja-Fernández and A. Tristán-Vega. Influence of noise correlation in multiple-coil statistical models with sum of squares reconstruction. *Magn Reson Med*, 67(2): 580–585, 2012.

S. Aja-Fernández, A. Tristán-Vega, and C. Alberola-López. Noise estimation in single- and multiple-coil magnetic resonance data based on statistical models. *Magn Reson imaging*, 27(10):1397–409, 2009.

S. Aja-Fernández, A. Tristán-Vega, and W.S. Hoge. Statistical noise analysis in GRAPPA using a parametrized noncentral Chi approximation model. *Magn Reson Med*, 65(4):1195–1206, 2011.

L.-C. Chang, G.K. Rohde, and C. Pierpaoli. An automatic method for estimating noise-induced signal variance in magnitude-reconstructed magnetic resonance images. In *Proceedings of SPIE Medical imaging*, volume 5747, pages 1136–1142, 2005.

P. Coupé, J.V. Manjón, E. Gedamu, D. Arnold, M. Robles, and D.L. Collins. Robust Rician noise estimation for MR images. *Med Image Anal*, 14(4):483–93, 2010.

T. Dhollander, J. Veraart, W. Van Hecke, F. Maes, S. Sunaert, J. Sijbers, and P. Suetens. Feasibility and advantages of diffusion weighted imaging atlas construction in Q-space. In *Proceedings of the 14th international conference on Medical image computing and computer-assisted intervention - Volume Part II*, MICCAI'11, pages 166–173, Berlin, Heidelberg, 2011. Springer-Verlag.

D.L. Donoho. De-noising by soft-thresholding. *IEEE Trans Inform Theory*, 41(3): 613–627, 1995.

J.X. Jim, S.B. Jong, and R.D. Swati. PULSAR: A matlab toolbox for parallel magnetic resoance imaging using array coils and multiple channel receivers. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering*, 31B (1):24–36, 2007.

C.G. Koay and P.J. Basser. Analytically exact correction scheme for signal extraction from noisy magnitude MR signals. *J Magn Reson*, 179(2):317–22, April 2006.

C.G. Koay, E. Ozarslan, and P.J. Basser. A signal transformational framework for breaking the noise floor and its applications in MRI. *J Magn Reson Imaging*, 197(2):108 – 119, 2009a.

C.G. Koay, E. Ozarslan, and C. Pierpaoli. Probabilistic Identification and Estimation of Noise (PIESNO): A self-consistent approach and its applications in MRI. *J Magn Reson*, 199(1):94 – 103, 2009b.

B.A. Landman, P.-L. Bazin, and J.L. Prince. Estimation and application of spatially variable noise fields in diffusion tensor imaging. *Magn Reson Imaging*, 27(6):741 – 751, 2009.

A Leemans, B Jeurissen, J Sijbers, and Derek K Jones. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine*, page 3537, Honolulu, USA, 2009.

J.V. Manjón, P. Coupé, L. Martí-Bonmatí, D.L. Collins, and M. Robles. Adaptive non-local means denoising of MR images with spatially varying noise levels. *J Magn Reson Imaging*, 31(1):192–203, 2010.

I.I. Maximov, E. Farrher, F. Grinberg, and N. J. Shah. Spatially variable Rician noise in magn reson imaging. *Med Image Anal*, 16(2):536 – 548, 2012.

K.P. Pruessmann, M. Weiger, M.B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magn Reson Med*, 42(5):952–962, 1999.

J. Rajan, D.H.J. Poot, J. Juntu, and J. Sijbers. Noise measurement from magnitude MRI using local estimates of variance and skewness. *Physics in Medicine and Biology*, 55(16):441–449, 2010.

P.M. Robson, A.K. Grant, A. J. Madhuranthakam, R. Lattanzi, D.K. Sodickson, and C.A. McKenzie. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. *Magn Reson Med*, 60(4):895–907, 2008.

J. Sijbers, A.J. den Dekker, J. Van Audekerke, M. Verhoye, and D. Van Dyck. Estimation of the noise in magnitude MR images. *Magn Reson Imaging*, 16(1): 87–90, 1998.

J. Sijbers, D.H.J. Poot, A.J. den Dekker, and W. Pintjens. Automatic estimation of the noise variance from the histogram of a magnetic resonance image. *Phys Med Biol*, 52:1335–1348, 2007.

J.-D. Tournier, S. Mori, and A. Leemans. Diffusion tensor imaging and beyond. *Magn Reson Med*, 65(6):1532–1556, 2011.

J. Wang, T. Kluge, M. Nittka, V. Jellus, N. Kuehn, and B. Kiefer. Parallel acquisition techniques with modified SENSE reconstruction (mSENSE). In *Proceedings of the First Würzburg Workshop on Parallel Imaging Basics and Clinical Applications*, page 92, Würzburg, 2001.

CHAPTER 8

Constrained parameter estimation

Contents

The work in this chapter has been published as:
J. Veraart, W. Van Hecke, and J. Sijbers. Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a Rician noise model. *Magnetic Resonance in Medicine*, 66(3):678-686, 2011.

8.1 Introduction

The least squares and maximum likelihood approaches for diffusion parameter estimation ignore a priori information about the feasible values of the estimated parameters. As diffusion of water molecules is a physical property of the tissue being measured, diffusional tensor estimates must be physically and biologically meaningful. Unfortunately, in many cases, diffusion and kurtosis values might lie outside a physically acceptable range due to the presence of noise, imaging artifacts such as Gibbs ringing, or misalignment of the diffusion-weighted images. In this chapter, we will list the constraints that should be imposed during the estimation of the DKI model parameters. For constraining the DTI fit, we refer to Koay and Basser [2006]

8.2 Linear inequality constraints

8.2.1 Constraint #1: lower bound constraint on $\mathrm{D_{APP}}$

To reconcile with the physical phenomenon of molecular diffusion, $\mathrm{D_{APP}}$ should be positive along each possible gradient direction. Alternatively, one can state that the diffusion tensor $\boldsymbol{D}$ should be positive definite. Its corresponding positive eigenvalues will result in FA values ranged between 0 and 1, the theoretically expected range. In practice, however, hyperintense voxels do show up in FA maps that are computed with unconstrained estimators. Indeed, since diffusion-weighted signals might exceed the nondiffusion-weighted reference values due to the presence of noise and imaging artifacts, those conventional estimators might results in diffusion tensors with negative eigenvalues. Typically, the radial eigenvalues are more likely to take negative values. Because of possibly low diffusion coefficients in the radial direction, noise or Gibbs ringing might dominate the expected signal attenuation. Early brute force strategies to enforce positive definiteness of the diffusion tensors included the rejection or reduction of the diffusion-weighted signals exceeding the nondiffusion-weighted reference signal [Koay et al., 2006]. Alternatively, negative eigenvalues were set to zero prior to calculation of the diffusion tensor parameters. Koay et al. [2006] proposed a more elegant constrained estimation strategy by representing the diffusion tensor by its Cholesky decomposition:

$$\boldsymbol{D} = \boldsymbol{C}^T \boldsymbol{C}, \tag{8.1}$$

with $\boldsymbol{C}$ an upper triangular matrix with non-zero diagonal elements:

$$\boldsymbol{C} = \begin{bmatrix} C_1 & C_2 & C_3 \\ 0 & C_4 & C_5 \\ 0 & 0 & C_6 \end{bmatrix}, \tag{8.2}$$

All diffusion estimators can now be reformulated in terms of the parameter vector:

$$\boldsymbol{\rho} = [C_1, C_2, C_3, C_4, C_5, C_6]^T . \tag{8.3}$$

For example, the NLS can be written as:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\rho}} \| \tilde{\boldsymbol{S}} - \exp\left(\boldsymbol{B}\boldsymbol{\beta}(\boldsymbol{\rho})\right) \|_2^2, \tag{8.4}$$

with

$$\boldsymbol{\beta}(\boldsymbol{\rho}) = \left[C_1^2, C_1C_2, C_1C_3, C_2^2C_4^2, C_2C_3 + C_4C_5, C_3^2 + C_5^2 + C_6^2\right]^T. \tag{8.5}$$

For the ML and (W)LLS estimators the same parameter substitution can be done to guarantee positive definite diffusion tensors. However, note that the object function is nonlinear in terms of ρ. So, no closed-form solution exists and iterative algorithms need to be applied to solve the optimization problem.

8.2.2 Constraint #2: lower bound constraint on $\mathrm{K_{APP}}$

There's no doubt that there should be a lower bound on $\mathrm{K_{APP}}$. The theoretical minimal kurtosis value equals -2 [Evans et al., 2000]. However, in the context of dMRI, -3/7 might be a more appropriate lower bound on $\mathrm{K_{APP}}$ as stated by Jensen et al. [2005]. Indeed, Jensen et al. [2005] showed that given water molecules confined to spherical pores, the kurtosis converges to -3/7 with the diffusion time going to infinity. In that asymptotic case, diffusion would be fully restricted. However, fully restricted diffusion is not expected in biological tissue. Indeed, the CHARMED model suggests the presence of both restricted and hindered compartments [Assaf and Basser, 2005]. Moreover, according to the CHARMED model, the population fraction of the restricted components does not exceed 50%. Therefore, it is generally assumed that $\mathrm{K_{APP}}$ should be positive. Empirical results substantiate that assumption [Jensen et al., 2005]. Let us recall Eq. (2.39):

$$\begin{aligned} K_{\mathrm{APP}}(\hat{\boldsymbol{g}}) &= \frac{\mathrm{MD}^2}{D_{\mathrm{APP}}(\hat{\boldsymbol{g}})^2} \sum_{i,j,k,l=1}^{3} \hat{g}_i\hat{g}_j\hat{g}_k\hat{g}_l K_{ijkl} \\ &= \frac{1}{D_{\mathrm{APP}}(\hat{\boldsymbol{g}})^2} \sum_{i,j,k,l=1}^{3} \hat{g}_i\hat{g}_j\hat{g}_k\hat{g}_l \tilde{K}_{ijkl}, \end{aligned} \tag{8.6}$$

with $\tilde{K}_{ijkl} = \mathrm{MD}^2 K_{ijkl}$, the modified kurtosis tensor owing to the linearized DKI model. As MD^2 and $\mathrm{D_{APP}}(\hat{\boldsymbol{g}})^2$ are always positive, $\mathrm{K_{APP}}(\hat{\boldsymbol{g}})$ is positive if

$$\sum_{i,j,k,l=1}^{3} \hat{g}_i\hat{g}_j\hat{g}_k\hat{g}_l K_{ijkl} > 0, \tag{8.7}$$

or

$$\sum_{i,j,k,l=1}^{3} \hat{g}_i\hat{g}_j\hat{g}_k\hat{g}_l \tilde{K}_{ijkl} > 0. \tag{8.8}$$

As a result, a positive $\mathrm{K_{APP}}$ along each unit direction of a densely sampled sphere can be enforced by imposing a set of linear constraints [Tabesh et al., 2011, Veraart et al., 2011].

8.2.3 Constraint #3: upper bound constraint on $\mathrm{K_{APP}}$

Assuming that the log-transformed diffusion-weighted signal along a fixed gradient direction, $\ln S(b, \hat{\boldsymbol{g}})$, is a monotonically decreasing function of the b-value, then the upper bound on $\mathrm{K_{APP}}(\hat{\boldsymbol{g}})$ can be derived as:

$$\mathrm{K_{APP}}(\hat{\boldsymbol{g}}) \leq \frac{3}{D_{\mathrm{APP}}(\hat{\boldsymbol{g}})b}, \tag{8.9}$$

as a necessary condition for the validity of the DKI model [Lazar et al., 2008]. Only then, the first derivative of

$$\ln S(b, \hat{\boldsymbol{g}}) \approx \ln S(0) - \mathrm{D_{APP}}(\hat{\boldsymbol{g}})b + \frac{1}{6}\mathrm{D_{APP}}(\hat{\boldsymbol{g}})^2\mathrm{K_{APP}}(\hat{\boldsymbol{g}})b^2, \tag{8.10}$$

with respect to b is negative within the range of acquired b-values. Since $\mathrm{D_{APP}}(\hat{\boldsymbol{g}})^2$ is positive along each gradient direction, inequality in Eq. (8.9) can be reformulated:

$$\mathrm{K_{APP}}(\hat{\boldsymbol{g}})\mathrm{D_{APP}}(\hat{\boldsymbol{g}})^2 \leq \frac{3\mathrm{D_{APP}}(\hat{\boldsymbol{g}})}{b}, \tag{8.11}$$

or

$$\sum_{i,j,k,l=1}^{3} \hat{g}_i\hat{g}_j\hat{g}_k\hat{g}_l\tilde{K}_{ijkl} - \frac{3}{b}\sum_{i,j=1}^{3} \hat{g}_i\hat{g}_j D_{ij} \leq 0 \tag{8.12}$$

One can see that the latter is a linear combination of the parameters of the linearized DKI model. We enforce the constraint along the directions of a densely sampled sphere by imposing a set of linear constraints [Tabesh et al., 2011, Veraart et al., 2011].

8.2.4 Matrix notation

In general, a constrained diffusion parameter estimator can be written as:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} h(\boldsymbol{\beta}), \tag{8.13}$$

such that

$$\boldsymbol{A}\hat{\boldsymbol{\beta}} \leq \boldsymbol{0}, \tag{8.14}$$

with h an arbitrary function of parameter vector $\boldsymbol{\beta}$ and $\boldsymbol{A}$ a matrix representing the linear (in)equality constraints [Tabesh et al., 2011]. Matrix $\boldsymbol{A}$ is defined as:

$$\boldsymbol{A} = \begin{bmatrix} 0 & -\boldsymbol{A_D^{(1)}} & 0 \\ 0 & 0 & -\boldsymbol{A_K^{(2)}} \\ 0 & -\frac{3}{b_{\max}}\boldsymbol{A_D^{(3)}} & \boldsymbol{A_K^{(3)}} \end{bmatrix}, \tag{8.15}$$

with

$$\boldsymbol{A}_D^{(1)} = \begin{bmatrix} \hat{n}_{1x}^2 & 2\hat{n}_{1x}\hat{n}_{1y} & 2\hat{n}_{1x}\hat{n}_{1z} & \hat{n}_{1y}^2 & 2\hat{n}_{1y}\hat{n}_{1z} & \hat{n}_{1z}^2 \\ \hat{n}_{2x}^2 & 2\hat{n}_{2x}\hat{n}_{2y} & 2\hat{n}_{2x}\hat{n}_{2z} & \hat{n}_{2y}^2 & 2\hat{n}_{2y}\hat{n}_{2z} & \hat{n}_{2z}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{n}_{Nx}^2 & 2\hat{n}_{Nx}\hat{n}_{Ny} & 2\hat{n}_{Nx}\hat{n}_{Nz} & \hat{n}_{Ny}^2 & 2\hat{n}_{Ny}\hat{n}_{Nz} & \hat{n}_{Nz}^2 \end{bmatrix},$$

with $\hat{\boldsymbol{n}}_i = \left[\hat{n}_{i_x}, \hat{n}_{i_y}, \hat{n}_{i_z}\right]$ the unit-length vector denoting the i^{th} out of N directions of a densely sampled sphere. $\boldsymbol{A}_D^{(3)}$ is defined analogously. The i^{th} row of the $N \times 15$ matrix $\boldsymbol{A}_K^{(2)}$ is given by:

$$\Big[\hat{n}_{i_x}^4 \quad 4\hat{n}_{i_x}^3\hat{n}_{i_y} \quad 4\hat{n}_{i_x}^3\hat{n}_{i_z} \quad 6\hat{n}_{i_x}^2\hat{n}_{i_y}^2 \quad 12\hat{n}_{i_x}^2\hat{n}_{i_y}\hat{n}_{i_z} \quad 6\hat{n}_{i_x}^2\hat{n}_{i_z}^2 \quad 4\hat{n}_{i_x}\hat{n}_{i_y}^3 \quad \cdots$$
$$12\hat{n}_{i_x}\hat{n}_{i_y}^2\hat{n}_{i_z} \quad 12\hat{n}_{i_x}\hat{n}_{i_y}\hat{n}_{i_z}^2 \quad 4\hat{n}_{i_x}\hat{n}_{i_z}^3 \quad \hat{n}_{i_y}^4 \quad 4\hat{n}_{i_y}^3\hat{n}_{i_z} \quad 6\hat{n}_{i_y}^2\hat{n}_{i_z}^2 \quad 4\hat{n}_{i_y}\hat{n}_{i_z}^3 \quad \hat{n}_{i_z}^4\Big].$$

Again, $\boldsymbol{A}_K^{(3)}$ is defined analogously.

8.3 Constrained parameter estimation

8.3.1 Constrained linear least squares estimators

Minimizing the linear least squares object function subject to linear equality and inequality constraints (e.g. Eq. (8.14)) can be done with quadratic programming (QP) [Gill et al., 1981]. A solution is given by the active set method. Initially, the unconstrained closed-form solution is computed and checked on its feasibility. An estimate is called feasible if none of the constraints is violated. If so, this unconstrained estimate is the solution, i.e. the global optimum, of the constrained optimization problem. However, if at least one constraint is violated, a feasible initial guess will be computed using linear programming methods to initiate an iterative process. So, basically, the active set method is an iterative method that solves a sequence of equality-constrained quadratic subproblems during which the set of active constraints are iteratively updated. The final solution always satisfies all constraints.

8.3.2 Constrained nonlinear estimators

The constrained nonlinear least squares (ordinary and conditional) functions – as well as the maximum likelihood function – form a well-behaved, twice differentiable, constrained nonlinear programming problem that can be efficiently solved by sequental quadratic programming (SQP) [Gill et al., 1981]. The optimizer requires the gradients of both the objective and the constraint functions. As they can both analytically be computed, the SQP algorithm is much faster at converging than the heuristic Nelder-Mead scheme we initially used [Nelder and Mead, 1965, Veraart et al., 2011].

8.4 Experiments

8.4.1 Single voxel simulations

Initially, Monte Carlo simulations (50000 trials) were done to evaluate the effect of varying SNR_{b_0}, i.e. the ratio between the noisefree nondiffusion-weighted signal and the noise level, on the MSE of the constrained and unconstrained least squares approaches for estimating FA, MD, and MK. Simulated Rice distributed data

with varying $\mathrm{SNR}_{b_0}^{-1}$ within a range of [10, 30] was simulated using Eq. (5.1). The ground truth diffusional tensors $\boldsymbol{D}$ and $\boldsymbol{W}$ were in-line values observed in the corpus callosum (cc) [Lätt et al., 2013]. The FA, MD, and MK were defined as $0.95 \times 10^{-3}\,\mathrm{mm}^2/\mathrm{s}$, 0.75, and 1.05, respectively. Diffusion-weighted signals with $b = 1000\,\mathrm{s/mm}^2$ and $b = 2500\,\mathrm{s/mm}^2$ were sampled along 60 directions, isotropically distributed over a unit sphere using Coulomb's law of repulsion [Jones et al., 1999]. Five nondiffusion-weighted signals were added. Rice distributed diffusion-weighted data were obtained by adding zero-mean complex Gaussian noise to the noise-free diffusion-weighted signals and calculating their magnitudes afterwards.

8.4.2 Whole brain simulations

Rice distributed simulation data sets, representing the whole human brain white matter were used for comparing the mean squared error (MSE) of the diffusion parameters obtained by the unconstrained and constrained least squares estimators. The simulated data sets were constructed as follows. First, ground truth tensors were obtained by voxel-wise fitting the DKI model to a real data set (see section 8.4.3: Data set 1). Second, a set of noise-free diffusion-weighted signals was reconstructed from those diffusion tensors using the DTI and DKI model for the DTI and DKI analyses, respectively. Gradient directions and b-values were in agreement with the single-voxel experiments. Third, 5000 sets of noisy diffusion-weighted signals with a uniform SNR_{b_0} of 20 were generated by adding 5000 realizations of complex Gaussian noise to the noise-free diffusion-weighted signals. From each set, the diffusion tensors were estimated. From the 5000 trials, the MSE in the estimation of FA, MD, and MK of the different estimators – constrained as well as unconstrained – were evaluated.

8.4.3 Real data experiments

The unconstrained WLLS, NLS and CLS estimator were run on two human data sets and the violations of either of the constraints was voxelwise examined. The following diffusion-weighted data sets of different healthy volunteers were acquired:

Data set 1: A first diffusion-weighted data set was collected on a 3T Philips Achieva MR scanner, using a 8-channel receiver head coil. Diffusion sensitizing was applied along 60 isotropically distributed gradient directions with $b = 1200\,\mathrm{s/mm}^2$ as well as $b = 2500\,\mathrm{s/mm}^2$. Additionally, one image without diffusion sensitization was acquired. Other imaging parameters were: TR/TE : 10265/107 ms; in-plane resolution: $1.75 \times 1.75\,\mathrm{mm}^2$; NEX: 1; slice thickness: 2 mm; axial slices: 70; and parallel imaging: SENSE with acceleration factor 2.

Data set 2: A third diffusion-weighted data set was acquired on a Siemens Trio (3T) MR scanner, using a 12-channel receiver head coil. Diffusion weighting was applied along 60 isotropically distributed gradient directions with $b = 1000\,\mathrm{s/mm}^2$ as well as $b = 2500\,\mathrm{s/mm}^2$. Additionally, 10 nondiffusion-weighted images were acquired. Other imaging parameters were: TR/TE : 6100/118 ms, in-plane resolution: $2.5 \times 2.5\,\mathrm{mm}^2$; NEX: 1; slice thickness:

2.5 mm; axial slices: 40; and parallel imaging: mSENSE with acceleration factor 2.

The diffusion-weighted data were corrected for motion and eddy currents, including signal modulation and b-matrix rotation [Leemans et al., 2009, Leemans and Jones, 2009]. For each voxel, constraints #2 and #3 were evaluated along 256 directions that were isotropically distributed on a sphere. For constraint #1, the eigenvalues were evaluated to determine constraint violations. Within white matter (WM) and gray matter (GM) – segmented using FSL's FAST algorithm [Zhang et al., 2001] – the percentage of voxels for which the constraints are not satisfied were computed and compared across both estimation strategies.

8.5 Results

8.5.1 Single voxel simulations

In Fig. 8.1, the constrained and unconstrained estimators are compared in terms of MSE as a function of $\mathrm{SNR}_{b_0}^{-1}$. The graphs indicate that the constrained estimation methods always outperform their unconstrained counterparts in terms of MSE in the estimation of MD, FA, and MK. The difference in performance increase with $\mathrm{SNR}_{b_0}^{-1}$.

8.5.2 Whole brain simulations

In Fig. 8.2, scatter plots show the relationship between the MSE in the estimation of the diffusion model parameters of WLLS, NLS, and CLS against that of their constrained counterparts, respectively. Each point in the scatter plot corresponds to a single voxel of the simulated data set. The MSE in the estimation of FA, MD and MK was significantly lower if the constraints were imposed during model fitting. During these simulations, statistical significance ($p < 0.01$) was shown with a paired Wilcoxon signed rank test.

8.5.3 Real data

In Fig. 8.3 and Fig. 8.4, MK and FA maps of the same axial slice, computed with unconstrained as well as the constrained estimators, are shown for data set 1 and 2, respectively. In the parameter maps estimated with the unconstrained algorithms (Fig. 8.3(a-c, g-i) and Fig. 8.4(a-c, g-i)), one can visually detect several outliers, which are related to constraint violations. On the one hand, the MK map is covered with black voxels, while, on the other hand, hyper intense FA values, i.e. FA > 1, can be observed . The former indicates negative kurtosis values (constraint #2); the latter indicates negative eigenvalues of the estimated diffusion tensor (constraint #1).

An overview of the spatial locations of constraint violations using unconstrained WLLS estimators is given in Fig. 8.5. Similar maps were computed for unconstrained NLS and CLS. However, the maps are not shown because of the high mutual correspondence. Violations of constraint #1 are shown in Fig. 8.5(a, d). Most of

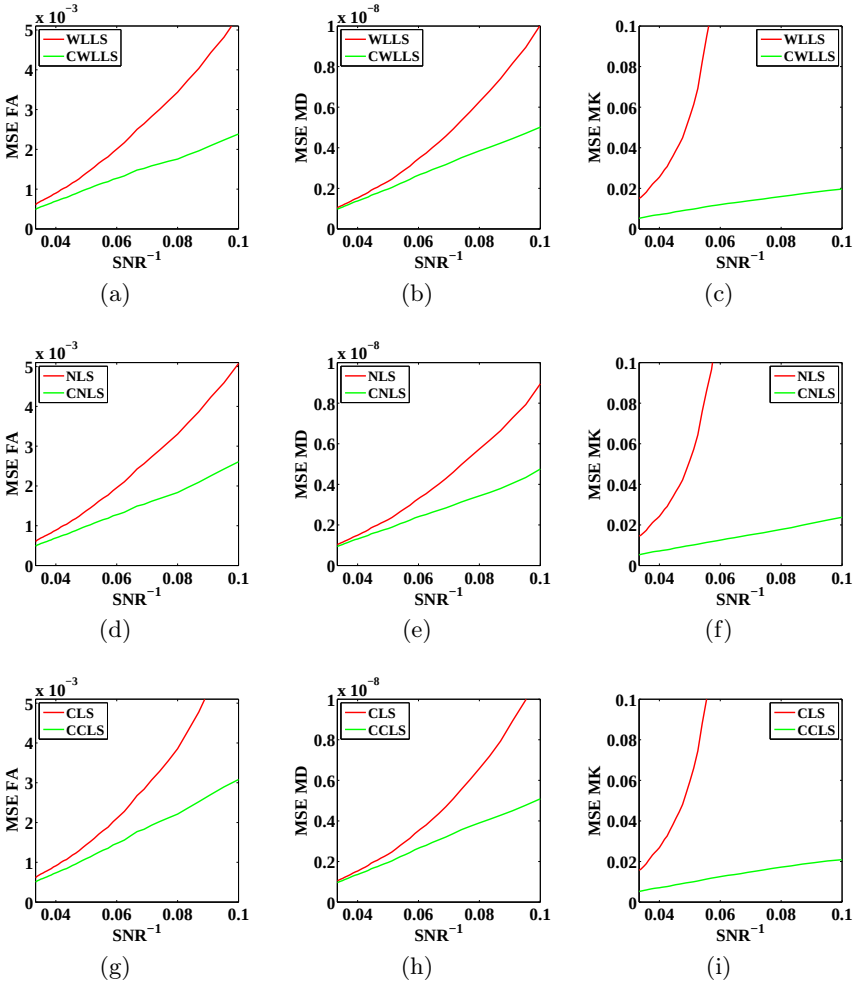
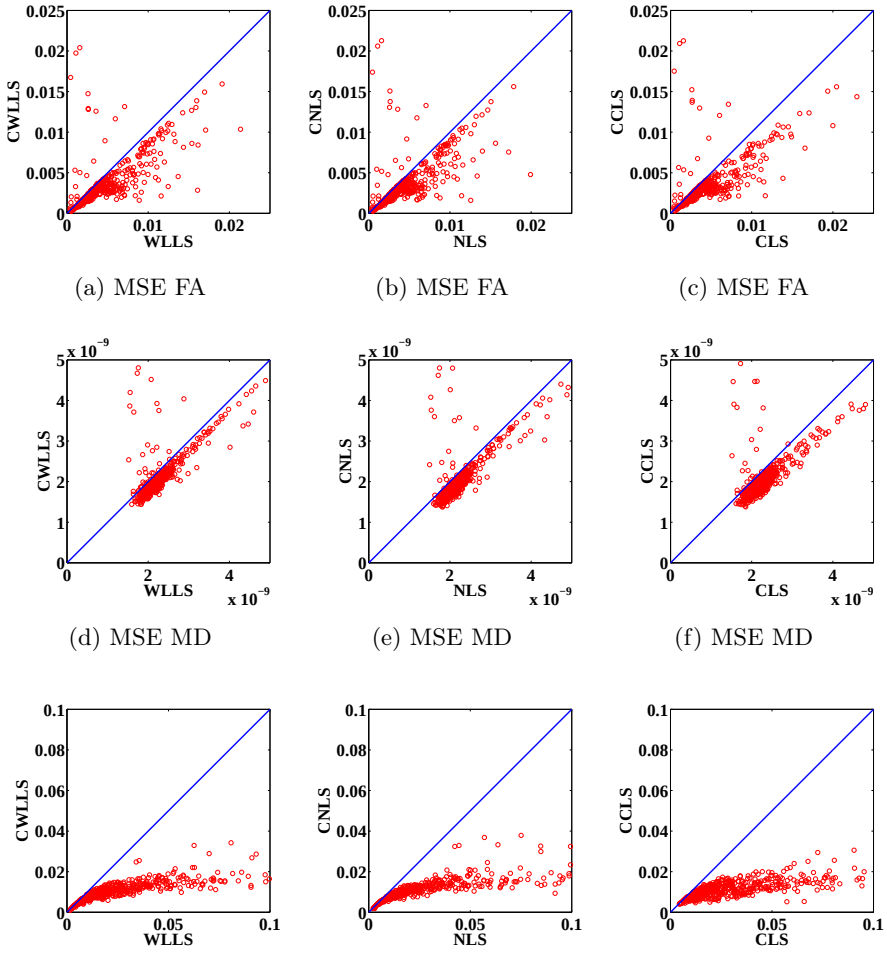


Figure 8.1: MSE in the estimation of FA, MD, and MK (g-i) are shown as a function of the $\mathrm{SNR}_{b_0}^{-1}$ for the different constrained and unconstrained least squares estimators.

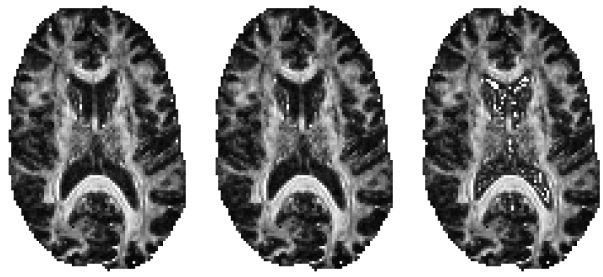


(a) MSE FA (b) MSE FA (c) MSE FA

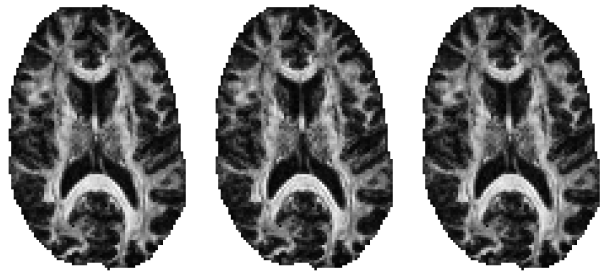
(d) MSE MD (e) MSE MD (f) MSE MD

(g) MSE MK (h) MSE MK (i) MSE MK

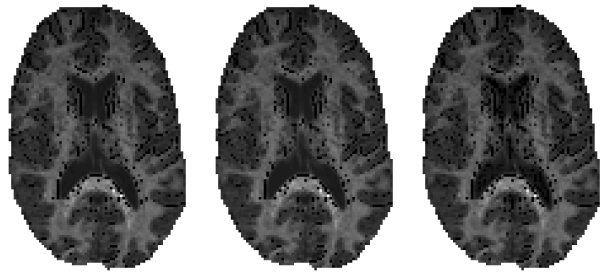
Figure 8.2: Scatter plots show the relationship between the MSE in the estimation of the diffusion model parameters of constrained against that of unconstrained parameter estimation. The blue lines are unit-slope-lines.



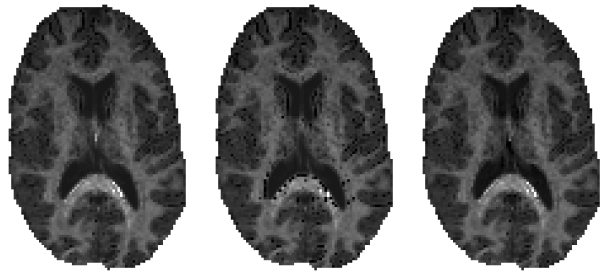
(a) WLLS (b) NLS (c) CLS



(d) CWLLS (e) CNLS (f) CCLS

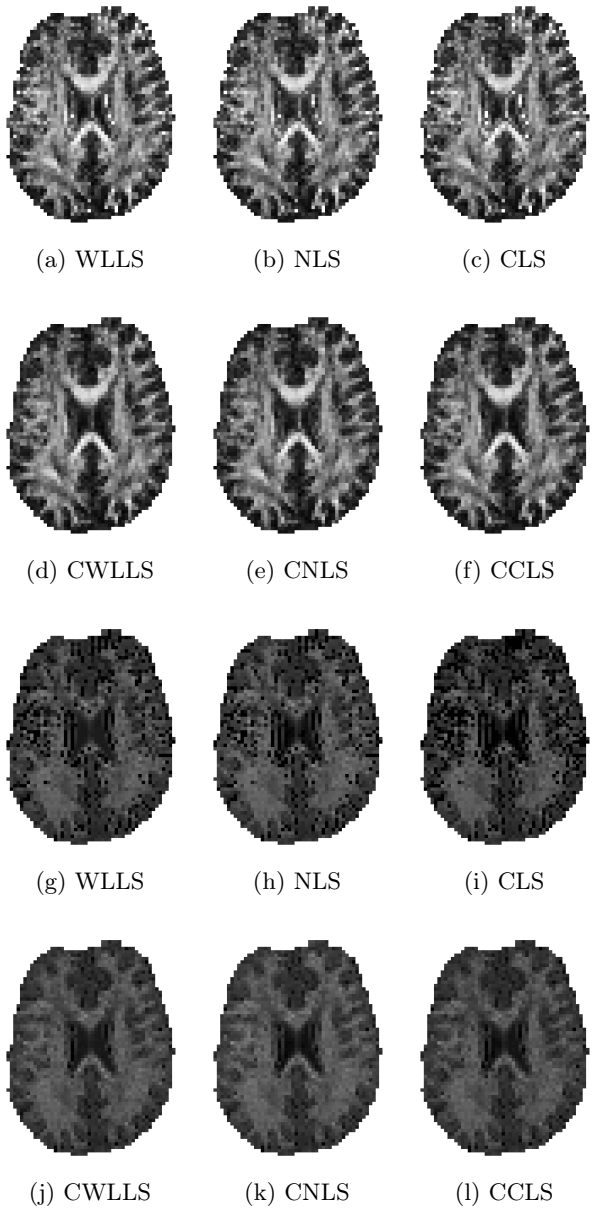


(g) WLLS (h) NLS (i) CLS



(j) CWLLS (k) CNLS (l) CCLS

Figure 8.3: FA, scaled between 0 and 1, (a-f) and MK, scaled between 0 and 1.5, (g-l) maps of the same axial slice, computed with unconstrained (a-c, g-i) as well as the constrained (d-f, j-l) estimators, are shown for real data set 1.



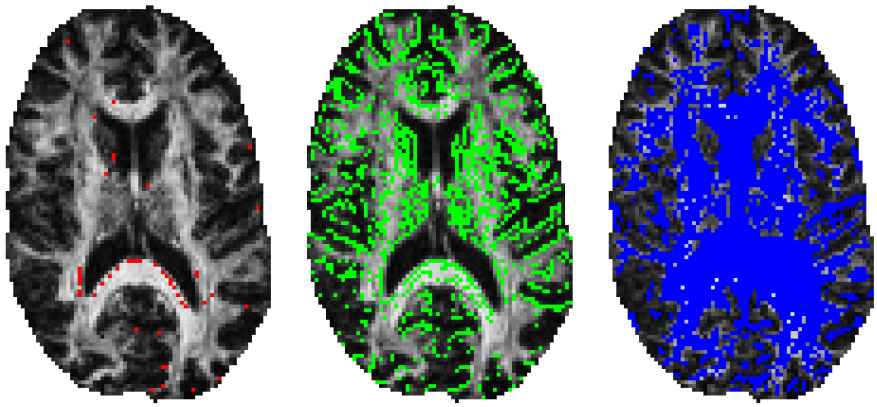
(a) WLLS (b) NLS (c) CLS

(d) CWLLS (e) CNLS (f) CCLS

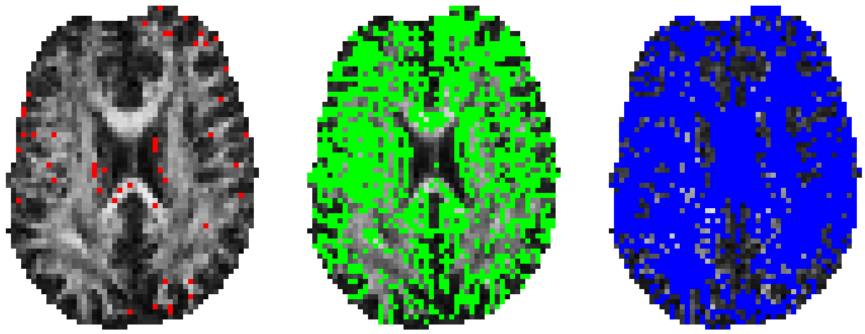
(g) WLLS (h) NLS (i) CLS

(j) CWLLS (k) CNLS (l) CCLS

Figure 8.4: FA, scaled between 0 and 1, (a-f) and MK, scaled between 0 and 1.5, (g-l) maps of the same axial slice, computed with unconstrained (a-c, g-i) as well as the constrained (d-f, j-l) estimators, are shown for real data set 2.

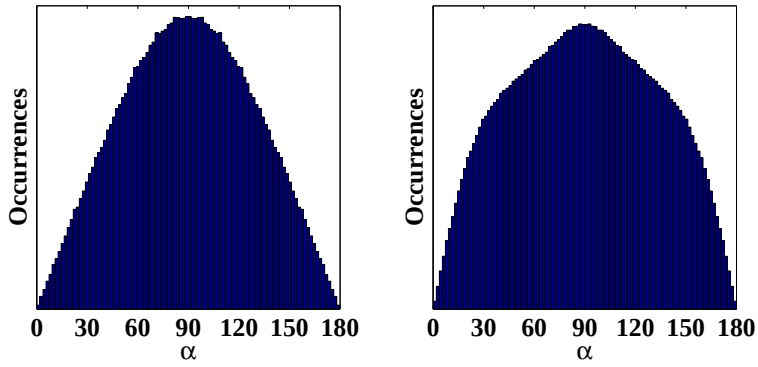


(a) Cosntraint #1 (b) Cosntraint #2 (c) Cosntraint #3



(d) Cosntraint #1 (e) Cosntraint #2 (f) Cosntraint #3

Figure 8.5: Spatial location of violations of constraint #1 (a,d), #2 (b,e), and #3 (c,f) for real data set 1 (a-c) and real data set 2 (d-f).



(a) Constraint #2 (b) Constraint #2

Figure 8.6: Directional dependency of prevalence of constraint #2 (a) and #3 (b) violations. The angle α is the angle in degrees between the direction along which the constraint was violated and the principal eigenvector the diffusion tensor. So, direction in the equatorial plane have $\alpha = 90°$. The histograms were obtained from both real data sets.

the violations appeared in the deep WM structures such as the genu and splenum of the cc. In Fig. 8.5(b, e), the voxels for which $K_{APP} < 0$ (constraint # 2 violations) were observed in at least one diffusion-weighting gradient direction are colored. Most voxels violating constraint #2 are within deep WM structures such as the cc. Violations of constraint #3 are shown in Fig. 8.5(c, f). Note that in numerous voxels, this condition was not satisfied. The DKI model resulted in too high kurtosis values within many voxels of various WM regions. Moreover, we observed a directional dependence of the prevalence. Indeed, constraints violations most often occur along radial directions. The high kurtosis and the low signal attenuation along those directions contribute to this observation. A more quantitative overview of the number of constraint violations is given in Table 8.1. The percentages of voxels violating constraint #1, #2 or #3 within WM and GM are tabulated. Note that positive definiteness on the diffusion tensors was generally satisfied when estimating the diffusion tensors with the DKI model. In less than 1.2% of the WM voxels, the diffusion tensor showed negative eigenvalues. The constraints on the kurtosis tensor, however, were violated in high percentages of the voxels within each tissue class. The most violated constraint was clearly constraint #3, the upper bound on the K_{APP}. Indeed, in more than 50% of the WM voxels the estimated tensors did not satisfy the constraint. Negative kurtosis values were observed in $> 30\%$ and $> 55\%$ of the WM voxels for Data set 1 and 2, respectively. More than 75% of the WM voxels showed constraints violations and in more than 35% of the GM voxels, physically irrelevant diffusional tensors were estimated.

8.6 Discussion

The simulation experiment as well as a real data study demonstrated that unconstrained parameter estimators not always result in physically and biologically

Table 8.1: The percentages of voxels violating constraint #1, #2 or #3 within WM and GM, estimated with the unconstrained WLLS, NLS, and CLS estimator.

	Data set 1					
	WLLS		NLS		CLS	
Constraint	WM	GM	WM	GM	WM	GM
#1	0.56	0	0.58	0	0.69	0.19
#2	33.62	22.56	32.47	20.91	37.30	25.83
#3	54.16	17.90	54.75	18.30	52.43	17.44
#1 ∪ #2 ∪ #3	75.83	39.23	75.58	38.07	76.06	40.93
	Data set 2					
	WLLS		NLS		CLS	
Constraint	WM	GM	WM	GM	WM	GM
#1	1.11	3.22	1.00	2.99	1.59	4.15
#2	61.84	55.41	57.10	50.00	72.90	66.74
#3	63.59	46.25	65.01	47.79	61.36	45.28
#1 ∪ #2 ∪ #3	91.37	80.11	90.34	78.00	92.79	84.01

relevant tensor estimates as the diffusion model is fitted to noisy DWIs, often corrupted with imaging artifacts. Therefore, diffusion and kurtosis parameters might be inaccurate and unreliable, hampering statistical analyses in clinical studies. Hence, constrained estimators are crucial in DKI analyses.

First, negative mean and directional kurtosis values were observed when using one of the unconstrained estimators. Typically, negative kurtosis values are observed in the deep WM structures, such as the cc. The genu and splenium of the cc as well as the simulated diffusion-weighted data, are characterized by a low radial diffusivity. Along low diffusivity directions, the noisy diffusion-weighted signal might appear being a concave function of the b-value, yielding negative estimated of directional kurtosis. In some extreme cases, the measured diffusion-weighted signal intensity may even exceed the nondiffusion-weighted signal intensity. As a result, negative estimate of diffusivity (violations of constraint #1) and extremely negative kurtosis values (violations of constraint #2) arises. The latter can be observed as black voxels in Fig. 8.4(g-i) and Fig. 8.4(g-i), while negative diffusivity yields hyper intense FA values. Obviously, negative kurtosis values did also appear within regions or along gradient directions with low kurtosis due to the variance of the selected estimator. Indeed, many constraint violations (#1, #2, and #3) are due to the expected variation of the estimator. Therefore, imposing the constraints will affect the estimators properties such as accuracy or precision.

Improving the fidelity of DKI estimators has initially been studied by Ardekani et al. [2010]. They imposed positive-definiteness on the diffusion kurtosis tensor by rewriting the 4^{th} order, 3D diffusion kurtosis tensor as a 2^{nd} order, 6D symmetric tensor with 15 unique elements. The Cholesky parameterization enables imposing positivity on the diffusion tensor as well as on the diffusion kurtosis tensor to obtain positive K_{APP} along each direction. The method, however, lacks an upper bound on the kurtosis values as a result of which a monotonically decreasing DKI model function is not guaranteed. As demonstrated in this study, the most often violated

constraint (constraint #3) is not imposed in their strategy. However, Tabesh et al. [2011] showed that by substituting $\mathrm{MD}^2\boldsymbol{K}$ by $\tilde{K}$, the linear least squares estimation problem becomes a special case of convex quadratic programming. Moreover, after substitution, all constraints on the model parameters are linear in terms of the model parameters as well.

In this chapter, we described how the approach introduced by Tabesh et al. [2011] can be adopted to constrain estimators with better performance than LLS in terms of accuracy and/or precision, i.e. the WLLS, NLS, CLS, and MLE. For a more detailed discussion of the constrained MLE, we would like to refer to Veraart et al. [2011]. Simulation and real data experiments indicated that the constrained estimation methods always outperform their unconstrained counterparts in terms of MSE in the estimation of MD, FA, and MK.

Bibliography

B.A. Ardekani, A. Tabesh, J.H. Jensen, J.A. Helpern, A. Bachman, and H. Kushner. An improved method for diffusional kurtosis estimation. In *Proceedings of the 18th Annual Meeting of the International Society for Magnc Reson Med*, page 1591, 2010.

Y. Assaf and P.J. Basser. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage*, 27(1):48–58, 2005. ISSN 1053-8119.

M. Evans, N. Hastings, and B. Peacock. Bernoulli distribution. In *Statistical Distributions*, chapter 4, pages 31–33. Wiley, New York, third edition, 2000.

P.E. Gill, W. Murray, and Wright M.H. *Practical optimization*. Academic Press, 1981.

J.H. Jensen, J.A. Helpern, A. Ramani, H. Lu, and K. Kaczynski. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magn reson imaging. *Magn Reson Med*, 53(6):1432–40, June 2005.

D.K. Jones, M.A. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magn reson imaging. *Magn Reson Med*, 42 (3):515–525, 1999.

C.G. Koay and P.J. Basser. Analytically exact correction scheme for signal extraction from noisy magnitude MR signals. *J Magn Reson*, 179(2):317–22, April 2006.

C.G. Koay, J.D. Carew, A.L. Alexander, P.J. Basser, and M.E. Meyerand. Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging. *Magnc Reson Med*, 55(4):930–6, April 2006. ISSN 0740-3194.

J. Lätt, M. Nilsson, R. Wirestam, F. Stahlberg, N. Karlsson, M. Johansson, P. C. Sundgren, and D. van Westen. Regional values of diffusional kurtosis estimates in the healthy brain. *J Magn Reson Imaging*, 37(3):610–8, March 2013. ISSN 1522-2586.

M. Lazar, J.H. Jensen, L. Xuan, and J.A Helpern. Estimation of the orientation distribution function from diffusional kurtosis imaging. *Magnc Reson Med*, 60 (4):774–81, October 2008. ISSN 1522-2594.

A Leemans and D K Jones. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn Reson Med*, 61(6):1336–1349, 2009.

A Leemans, B Jeurissen, J Sijbers, and Derek K Jones. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine*, page 3537, Honolulu, USA, 2009.

J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

A. Tabesh, J.H. Jensen, B.A. Ardekani, and J.A. Helpern. Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. *Magn Reson Med*, 65(3):823–36, 2011.

J. Veraart, W. Van Hecke, and J. Sijbers. Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a Rician noise model. *Magn Reson Med*, 66(3):678–686, 2011.

Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Trans Medical Imaging*, 20(1):45–57, 2001.

Part III

Atlas construction

CHAPTER 9

Construction of a DTI atlas of the Sprague Dawley rat brain

Contents

The work in this chapter has been published as:
J. Veraart, T.B. Leergaard, B.T. Antonsen, W. Van Hecke, I. Blockx, B. Jeurissen, Y. Jiang, A. Van der Linden, G.A. Johnson, M. Verhoye, and J. Sijbers. Population-averaged diffusion tensor imaging atlas of the Sprague Dawley rat brain. *NeuroImage*, 58(4):975-983, 2011

9.1 Introduction

Because of the rat's ability to accurately predict the human response, the rat has become the most widely studied experimental animal model for biomedical research, which might involve microsurgical techniques, serial sampling of the cerebrospinal fluid (CSF), and repetitive *in vivo* neuroimaging [Cenci et al., 2002, Cozzi et al., 2008]. The rat exhibits physiological characteristics similar to those of humans and therefore, several strains, substrains, and genetically modified rats are employed to mimic various neurological disorders such as Huntington Disease [von Hörsten et al., 2003], Amyotrophic lateral sclerosis [Vermeiren et al., 2006], and Alzheimer disease [Liu et al., 2008].

Several atlases of the rat brain have been developed and provide invaluable resources for a wide range of applications in various types of neuroscience studies, including visualization, identification, and precise localization of specific brain areas, stereotaxic surgery, delineation of brain regions of interest (ROI), and registration of information such as gene expression locations. To date, the most widely used rat brain atlases are the stereological, histology-derived atlases of the Wistar and Sprague Dawley (SD) rat brain by Paxinos and Watson [1986, 2007] and Swanson [1992, 1998, 2004], respectively. These atlases describe the brain in a series of two-dimensional diagrams, indicating the names and boundaries of areas and nuclei. Despite their widely use, the histology-based atlases have several limitations. First, histology-based atlases are inherently two-dimensional and restricted to standard section planes. Second, the procedures of specimen fixation, sectioning, and staining involved in histological processing can deform the brain shape, which might be a source of significant inaccurate localization of brain areas. During the last decade, several average *in vivo* templates of the rat brain and atlases of its major structures have been developed with magnetic resonance imaging (MRI), a non-invasive, three-dimensional, and *in vivo* imaging technique, which enables accurate identification of a large number of anatomical structures [Schweinhardt et al., 2003, Schwarz et al., 2006, Hess et al., 2005]. However, all proposed atlases lack orientational information about the white matter anatomy, which constitutes a complex network of axons connecting different brain regions. This is conclusively due to the homogeneous appearance of the white matter structures in conventional MRI and in histology preparations. The lack of clear anatomical boundaries hampers identification and delineation of specific white matter fiber bundles.

Magnetic resonance diffusion tensor imaging (DTI) is a widely explored, and exceptional modality for quantifying the random walks of water molecules in biological tissue [Basser et al., 1994]. Unlike other MR techniques, it offers the possibility of characterizing and visualizing the structural connectivity of distinct anatomical networks within the brain *in vivo* and non-invasively [Jones, 2008]. With DTI, the estimated displacement profile of the diffusing molecules can be interpreted as an ellipsoidal iso-probability surface - described by a diffusion tensor (DT). In the white matter, the first eigenvector of the DT, thus the preferred local displacement direction of the water molecules, coincides with the orientation of major fiber tracts, as a result of which DTI provides rich anatomical contrast highly suitable for accurate delineation of white matter regions and subregions [Lin et al., 2001, Kaufman et al., 2005, Dauguet et al., 2007].

To provide a resource for efficient three-dimensional parcellation and analysis of rat brain neuroimaging data, we here present an accurate three-dimensional atlas of the normal adult SD rat brain constructed from precise, manually delineated anatomical labels. Selected gray and white matter structures were manually delineated in a high-resolution, diffusion weighted and anatomical MRI scan obtained *ex vivo* with the brain *in situ* within the skull. To minimize bias due to sample-specific features in the *ex vivo* sample, we nonlinearly mapped the *ex vivo* atlas onto a population-averaged *in vivo* rat brain template. The construction of that population-averaged template allowed studying the anatomical variation within the sample population [Aggarwal et al., 2009]. The proposed atlas offers an average anatomical template for accurate identification of the white and gray matter structures, which could, for example, be used for ROI delineation of the rat brain imaging data.

9.2 Materials and Methods

9.2.1 Animal samples and preparation

Animal procedures were approved by the local institutional animal welfare committee at the Universities of Oslo and Antwerp, and were in compliance with National Institutes of Health and European Community guidelines for the use and care of laboratory animals. For the construction of the *in vivo* population-averaged template, nine inbred male SD aged ∼12 months were used, while for the *ex vivo* atlas, one inbred male SD rat aged ∼18 months was used. The animals were bred at the Franz-Penzoldt-Center, Experimental Therapy, Friedrich-Alexander-University of Erlangen-Nürnberg, Germany.

Before *ex vivo* MR imaging, the specimen was prepared for active staining according to [Johnson et al., 2002], with a contrast agent mixed with a fixative to enhance the MRI signal. Following a brief inhalation induction with 4% isoflurane (Abbott Laboratories, Illinois, USA), the animal was deeply anesthetized by intraperitoneal injection of sodium pentobarbital (50 mg/kg) and transcardially perfused with 120 ml of a mixture of 0.9% saline, ProHance® (10:1 v:v; gadoteridol, Bracco Diagnostics, Inc, Princeton, NJ), and Heparin (5000 IE units / ml; Leo Pharma A/S, Ballerup, Denmark), followed by 120 ml of freshly prepared 4% paraformaldehyde with ProHance (10:1, v:v). The animal head was isolated and stored in 0.9% saline with ProHance (10:1 v:v). During all procedures the brain was left *in situ* within the cranium to limit physical distortions.

9.2.2 In vivo MR imaging

Diffusion weighted images (DWIs) were acquired at the Bio-Imaging Lab on a 9.4T Bruker Biospec scanner (Ettlingen, Germany) using a fast spin echo sequence with an encoding scheme of 6 diffusion weighting gradient directions using $b = 800\,\mathrm{s/mm^2}$, TR/TE= 2200/34 ms, $\delta = 5\,\mathrm{ms}, \Delta = 18\,\mathrm{ms}$, acquisition matrix = 256×128 (zerofilled to 256×256), FOV= $35 \times 35\,\mathrm{mm^2}$, 28 slice with thickness 0.43 mm. Additionally, one image without diffusion weighting (b_0) was acquired. For each

animal, diffusion weighted datasets were acquired for 7 repetitions (two averages each), which resulted in a total scan time of $\sim$ 4h.

9.2.3 Ex vivo imaging

DWIs were acquired at the Duke Center for In Vivo Microscopy using a 7T Magnex 7.0 T/210 mm bore magnet controlled by GE EXCITE consoles. A Specimen was imaged in a solenoid rf coil fabricated from a continuous sheet of high frequency microwave substrate (Roger Corp, Rogers, Ct). A diffusion-weighted spin-echo pulse sequence with extended dynamic range [Johnson et al., 2007] was used to acquire 3D volume images ($FOV = 45 \times 22.5 \times 22.5\,\mathrm{mm}^3$, TR/TE= 100/15.6 ms, NEX=2). Diffusion encoding was performed using a pair of half-sine gradient pulses ($\delta = 3.2\,\mathrm{ms}/\Delta = 8.3\,\mathrm{ms}$), using $b = 800\,\mathrm{s/mm}^2$. A reduced encoding DTI methodology [Jiang et al., 2004] was employed, such that the dataset consisted of a fully encoded $512 \times 256 \times 256$ (readout $\times$ phase $\times$ slice) matrix-size b_0 (i.e., $b \approx 0$) and 12 reduced encoded ($512 \times 128 \times 128$) diffusion-weighted images sensitized in each of an optimized set of 12 directions [Papadakis et al., 1999]. Each reduced encoded diffusion-weighted image was reconstructed to $512 \times 256 \times 256$ matrix size by a corrected keyhole algorithm [Jiang and Hsu, 2005] with the b_0 image as the constraining reference, resulting in 88 μm isotropic resolution. The acquisition time for the complete DTI dataset was approximately 18h. A RF refocused spin echo image with the same FOV and resolution was acquired with TR = 50 ms, $TE = 5$ ms, and NEX=1. Active staining with Prohance reduces the T_1 of all the tissues to <100 ms so this sequence produces anatomical images similar to those one would obtain with proton density weighting in unstained tissues.

9.2.4 DTI data analysis

The DTI model is given by:

$$\ln S(b, \boldsymbol{\theta}) = \ln S_0 - b \sum_{i,j=1}^{3} g_i g_j D_{ij}. \tag{9.1}$$

In Eq. (9.1), g_i is the i^{th} component of diffusion weighted gradient direction $\boldsymbol{g}$ and b probes the diffusion weighting strength. $S(b)$ and S_0 are the diffusion weighted and non-diffusion weighted signal intensities, respectively. D_{ij} is the ij^{th} element of the fully symmetric apparent diffusion tensor $\boldsymbol{D}$. The DTI model is parameterized by $\boldsymbol{\theta}$, which includes 7 parameters: S_0 and 6 independent elements of $\boldsymbol{D}$, [D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{xz}, D_{yz}].

The parameter vector $\boldsymbol{\theta}$ was estimated voxelwise by fitting Eq. (9.1) to the natural logarithm of the diffusion weighted data - corrected for motion and eddy currents using the FSL toolbox [Jenkinson et al., 2002]- such that the sum of the weighted squared differences was minimized (see Chapter 5). Several diffusion parameters, such as the fractional anisotropy (FA), mean (MD), radial ($D_\perp$) and axial ($D_\parallel$) diffusivity, were calculated voxelwise by means of the eigenvalue decomposition of the related diffusion tensors [Le Bihan et al., 2001].

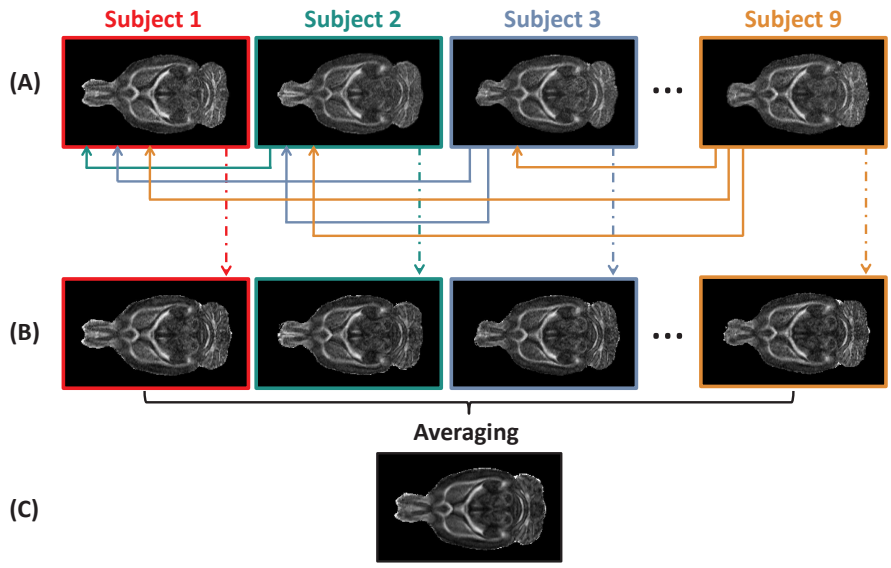
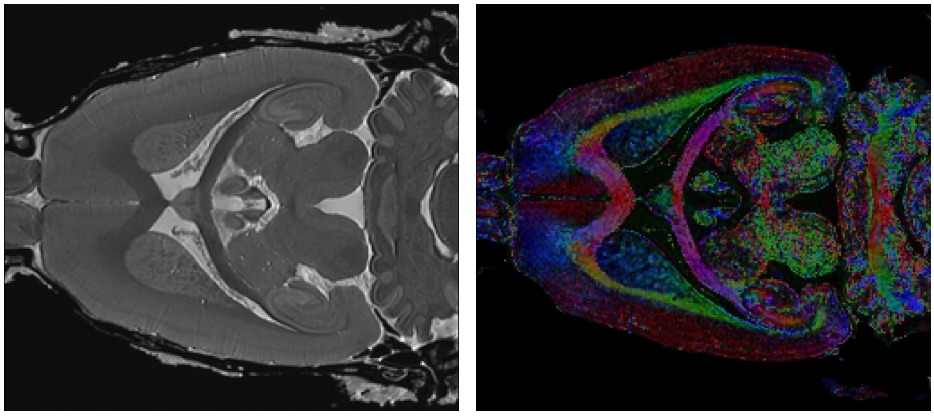


Figure 9.1: An overview of the *in vivo* population-averaged template construction. In (A), the FA maps of the subjects, affinely aligned to an arbitrarily chosen subject, are shown. The deformation field that maps subject j onto subject i is calculated and denoted as T_{ij}, with $i < j$, and $j = 1 \cdots 9$. The calculation of the deformation fields was done with a nonrigid registration algorithm based on a viscous fluid model. Next, for subject i, an average mean deformation field is computed as the average deformation to all other subjects: $T_i = \sum_{j=1}^{9} \boldsymbol{T}_{ji}$, with T_{ji} the inverse of T_{ij} . The average deformation fields, T_i, are applied to the corresponding DTI datasets (B). The average of the deformed DTI datasets resulted in the population based atlas (C). In (B) and (C), the DTI datasets are represented by the FA maps.

9.2.5 Population-averaged brain template

All nine *in vivo* diffusion weighted datasets were used to create a population-averaged DTI template that robustly preserves the orientational DT information and contains a minimal bias towards any specific individual dataset. First, all images were affinely aligned to a randomly chosen subject to correct for global misalignments. The affine registration was performed with in-house ITK software based on maximizing the mutual information between FA maps as they provide high white/grey matter contrast [Maes et al., 1997]. Next, for each dataset, a mean shape template, defined as a minimal deformation target (MDT) by [Kochunov et al., 2001], was constructed by nonrigidly transforming a single image in a way that the deformed image requires the least amount of deformation to all other images in the group. An identical MDT brain should be obtained regardless of the image from which it was constructed. Although, in practice, all MDTs were very similar, some unresolved residual variations, explained by the topology preserving property of the coregistration algorithm, were noticeable. Those variations were reduced by a voxel-wise averaging over all MDTs, resulting in a template that is called the population-averaged DTI template [see Fig. 9.1; Wang et al., 2005, Van



(a) *ex vivo* T_1 (b) *ex vivo* DEC

Figure 9.2: (a) A single horizontal slice of the *ex vivo* T_1 image, and (b) the direction encoded (DEC) FA map, with RGB-colors representing the orientation of the first eigenvector of the DTs and intensity values in proportion to the FA value, of a single *ex vivo* sample are shown. The red, green and blue color correspond to the mediolateral, dorsoventral, and anterioposterior orientations, respectively. These maps were used for the manual delineation of the anatomical structures.

Hecke et al., 2008]. [Wang et al., 2005, Van Hecke et al., 2008].

This used nonrigid coregistration algorithm computes a fluid-model based deformation field via voxel-by-voxel diffeomorphic mapping from a multichannel floating image to a multichannel reference image [D'Agostino et al., 2003]. Mutual information was used as a cost function. The calculation of the deformation field was steered by all unique diffusion tensor elements to take full advantage of the relevant information that was encoded in DTI data, particularly the tensor orientation, and thus to reduce local misalignments in the white matter tracts [Van Hecke et al., 2007]. Since all images acquired in this study included orientational information, tensor reorientation (PPD) and recalculation of the DWIs was performed after the deformation in order the preserve the alignment of the diffusion tensors and the underlying structures [Alexander et al., 2001].

9.2.6 Manual delineation of brain structures on *ex vivo* MR data

In the *ex vivo* DTI volume, a selection of white matter and gray matter regions were manually segmented on basis of T_1-weighted (Fig. 9.2a) and DTI contrast (FA and principal eigenvector orientation; Fig. 9.2b), following a stepwise procedure to utilize complementary information in the structural and diffusion data. Image processing and anatomical delineations were performed using the ITK-SNAP (version 1.6; [Yushkevich et al., 2006]; www.itksnap.org) and Amira® (Visage imaging, Inc., San Diego, CA) software packages. First, structures were delineated on basis of T_1 white/grey matter contrast observed in coronal, sagittal, and horizontal image slices

(Fig. 9.3(a,d)). Second, in regions where anatomical boundaries were ambiguous or invisible in T_1-weighted images, DTI maps were used to adjust boundaries (Fig. 9.3(b,e)). To this end, the orientation of the principal eigenvector was co-displayed with the T_1-images, in separate channels for anterioposterior (blue), mediolateral (red), and dorsoventral (green) orientations. The observed diffusion orientations were evaluated against anatomical landmarks visible in series of coronally and sagittally oriented histological sections (Fig. 9.3(c,f); T.B. Leergaard, A.M. Dale, and J.G. Bjaalie, unpublished work, see also [Leergaard et al., 2010]) stained for myelin following a standard procedure modified from [Woelche, 1942]. Two standard rat brain atlases were used as reference [Swanson, 2004, Paxinos and Watson, 2007], and additional predefined anatomical criteria (see below) were employed to close anatomical boundaries when these were not unequivocally visible in T_1 or DTI images. This particularly concerned locations where myelinated fibers pass between white matter fiber bundles and GM, the transition from hyperintensive to hypointensive T_1 contrast is gradual and boundaries ambiguous. The lateral ventricles and the associated ependymal layer were segmented from T_1 images [Swanson, 2004].

9.2.6.1 White matter

The corpus callosum (*cc*) consists of mediolaterally oriented commisural fibers that continue in the external capsule (*ec*) underlying the cerebral cortex (*Cx*) [e.g. Heimer et al., 1967, Sargon et al., 2003]. The *cc* and *ec* were here segmented as one structure. The boundary of the *ec* and the overlying *Cx* is ambiguous due to the high number of myelinated fibers passing between both structures (Fig. 9.3(a,d)). Here, DTI maps provided excellent contrast between coherent fibers oriented along the *ec*, and radial fiber orientations in the *Cx* (Fig. 9.3(b,e)). The cingulum (*cg*) bundle lies dorsomedial to the *ec*, and contains anterioposteriorly oriented limbic fibers [e.g. White, 1959, Swanson and Cowan, 1979]. The boundary between the *cg* and *ec* is indistinguishable in T_1 contrast (Fig. 9.3(a)), but readily segmented in DTI images showing anterioposterior and mediolateral orientations associated to the *cg* and *cc*/*ec*, respectively (Fig. 9.3(b,e)). Subcortically projecting corticofugal axons pass through the striatum (giving rise to its characteristic striated texture) and converge into the internal capsule (*ic*), which is continuous with the cerebral peduncle (*cp*) and pyramidal tract (*py*) [e.g. Coleman et al., 1997]. We here delineated the *ic* and *cp* as one structure. The *ic* was defined as a coherent region of white matter (i.e. a continuous cluster of voxels with hyperintensive T_1 contrast) located in between the globus pallidus (*GP*) and the thalamus, thus excluding the dispersed and apparently solitary fiber bundles within the striatum. Obviously, these discontinuities relate to the thickness of fiber bundles relative to the employed voxel size. The further trajectory of the *ic*/*cp* was readily delineated in T_1-weighted images to the level of the pons, where DTI images were used to differentiate the *cp* from the ascending medial lemniscus (*ml*). The anterior commissure (*ac*), which decussates anterior of the fornix (*f*) columns, has an anterior part (*aca*) extending anteriorly into the olfactory bulb (*OB*), and a posterior part (*acp*) extending laterally into the striatum. Both the *aca* and *acp* were readily identified in both T_1 and DTI images. The optic tract (*opt*) contains fibers from the optic nerves

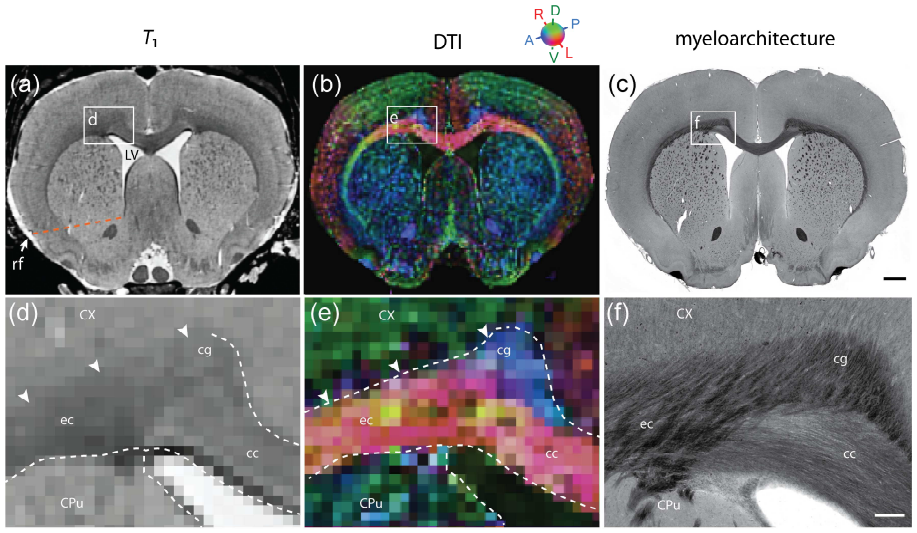


Figure 9.3: Coronal *ex vivo* T_1-weighted (a,d) and DEC-DTI (b,e) slices, shown together with corresponding images (c,f) of a coronal, myelin stained section from a different animal. White frames in (a-c) indicate the position of the enlarged images in (d-f). Images were manually segmented on basis of T_1, supplemented by DTI contrast in regions where T_1 contrast was insufficient. The detailed interpretation of DTI images was aided by inspection of myelin fiber orientations in corresponding histological section images. It is not possible to distinguish the cingulum (*cg*) and corpus callosum (*cc*) in the T_1 images, but the anterioposteriorly oriented fibers of the *cg* (e, f) are standing out in DT images (blue color in (b,e)). The dorsal boundary of the external capsule (*ec*) is ambiguous in T_1-weighted images (arrowheads in (d)), but readily identified mediolaterally oriented diffusion orientations in DTI (red color in (b,e)). The dotted line in (a) indicates the imaginary boundary between the dorsal caudate putamen (*CPu*) and the nucleus accumbens, drawn as a line between the rhinal fissure and ventral tip of the lateral ventricle. Scale bars, 1 mm and 250 μm.

and runs along the ventrolateral surface of the diencephalon from the optic chiasm (*och*) to the lateral geniculate nucleus of the thalamus. It is difficult to distinguish from neighboring fiber tracts in T_1 images, but detectable in DTI maps showing mediolateral and dorsoventrally oriented diffusion. The fimbria (fi) and fornix (f) of the hippocampus contain the main output fiber bundles of the hippocampus, and were mainly delineated from DTI images, aided by histological material showing the corresponding orientations of myelinated fibers. The fi forms a band of white matter along the lateral and rostral aspects of the hippocampus, and the f a distinct flat bundle close to the midline below the *cc* [Amaral and Lavenex, 2007]. The boundary of the fi towards the lateral ventricle (LV) and its ependymal lining was delineated in T_1 images.

9.2.6.2 Subcortical gray matter

Most parts of the basal ganglia (the striatum, the globus pallidus (GP), entopeduncular nucleus (EP), and the substantia nigra (SN)) were segmented on basis of the T_1-weighted images. The striatum includes the caudate-putamen (CPu) complex and the core of the nucleus accumbens (Acb) [Gerfen, 2004], and was dorsally delineated along the *ec*, and ventrally along the relatively hyperintensive contrast of the olfactory tubercle. Since the posterior limit of the striatum is ambiguous, it was arbitrarily set at the most anterior level where the distinct CA3 field of the hippocampus is visible in coronal slices. The striatum is usually divided into a dorsal (CPu) and ventral (Acb) region on basis of hodology and neurochemistry [Voorn et al., 2004]. As the boundary between the CPu and Acb is unclear in histological and tomographical material, it is usually defined by an imaginary line between the inferior tip of the LV and the rhinal fissure (rf) [Ingham et al., 1998, Van de Berg et al., 2000, Voorn et al., 2004]. In our material, the following subdivisions of the dorsal and ventral striatum were employed: Anterior of the decussation of the *ac*, a line was drawn between the rf to the inferior tip of the LV. In further anterior regions where the LV could not be distinguished in T_1 images, the ventromedial tip of the *cc* was used as a substitute landmark. Posterior of the decussation of the *ac* a line was drawn between the left and right rf.

9.2.7 Correction of *ex vivo* atlas

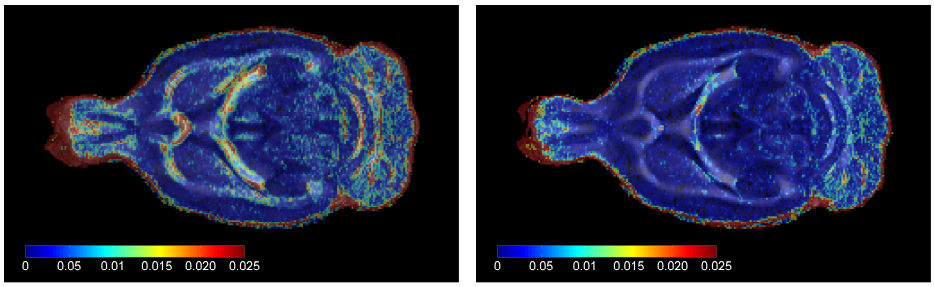
Because the *ex* vivo atlas is based on a single rat brain, it might be biased towards subject-specific anatomical features. To minimize these effects, we mapped the high-resolution, *ex vivo* images and parcellation map nonrigidly, i.e. using the viscous fluid model, into the *in vivo* population-averaged space.

9.3 Results

9.3.1 Population-averaged DTI template

9.3.1.1 Registration quality

The construction of the population-averaged was initialized by the affine alignment of each *in vivo* dataset to a single, arbitrarily chosen, reference *in vivo* dataset. After affine registration, however, there was still a considerable amount of misregistration noticeable as shown in Fig. 9.4(a). In the figure, the intersubject FA variance, voxelwise calculated across all affinely aligned subjects, is visualized on top of the corresponding average FA map. As shown in Fig. 9.4(b), the amount of misregistration has been significantly reduced by applying the second step of the construction of the population-averaged DTI brain template, which involved nonrigid deformation of the datasets. Fig. 9.4(b) shows the FA variance, calculated after warping each subject onto the population-averaged DTI template, on top of the population-averaged FA map. One may in particular appreciate the clearly decreased FA variance in all boundaries of the white matter structure, corresponding to a sharper white/grey matter contrast.



(a) (b)

Figure 9.4: (a) The average *in vivo* FA template (in gray color scale) of a single horizontal slice after affine alignment of all subjects to a single, arbitrarily chosen, subject, overlaid with the intersubject FA variance map (shown in spectral color scale). Blue regions indicate low variability, while red indicate high variability. (b) Compared to (a), the FA variance and average was calculated after warping each subject onto the population-averaged DTI template.

To further quantify and compare the quality of the different registration steps, we evaluated similarity between aligned DTI datasets - represented by their FA maps - using the normalized correlation coefficient. We calculated the similarity between pairs of DTI datasets, which were (I) not aligned (original data), (II) affinely aligned, (III) nonrigidly aligned to a single subject, and (IV) nonrigidly aligned to the population-averaged DTI brain template. Note that (I), (II), and (IV) were steps included in the construction of the population-averaged DTI brain template,

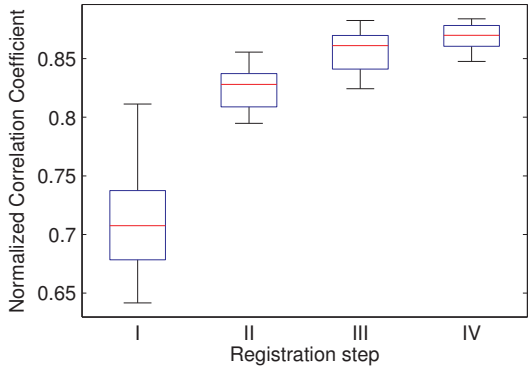


Figure 9.5: Similarity between pairs of DTI datasets (represented by the FA map), which were (I) not aligned (original data), (II) affinely aligned, (III) nonrigidly aligned to a single subject, and (IV) nonrigidly aligned to the population-averaged DTI brain template, was evaluated with the normalized correlation coefficient. Each set of FA maps resulted in 36 pairwise calculated normalized correlation coefficients, of which the median value is indicated by the red lines. The lower and upper edges of the boxes correspond to the 25^{th} and 75^{th} percentiles, respectively.

while (III) was only evaluated for comparison purposes. For each registration step, a FA map was generated for all nine individual subjects. Each set of FA maps resulted in 36 pairwise calculated normalized correlation coefficients, which could be used to compare the image alignment across the different registration steps (see Fig. 9.5). Adding a nonrigid coregistration step during atlas construction resulted in a significantly increase of pairwise normalized correlation coefficients. Furthermore, the construction of a population-averaged template (IV) yielded significantly higher correlation coefficients compared to the subject-based atlas construction approach (III). Statistical difference between the consecutive steps was demonstrated with a paired student t-test, with significance level set to 0.05.

9.3.1.2 Anatomical variability

To evaluate the degree of anatomical variability within the sample population, a deformation field was computed for each single subject by warping the population-averaged brain template to each subject using the nonrigid coregistration algorithm. The deformation fields, i.e. the length of the deformation vectors, quantify the anatomical differences among the individual brains and the average anatomy, represented by the population-averaged template in each voxel. By voxelwise averaging the length of the deformation vectors over all studied rats, an anatomical variability magnitude (AVM) map is obtained [Aggarwal et al., 2009]. The intensity levels in the map denote distances in millimeters, and represent the average anatomical variability in the adult rat brain across the sample population (see Fig. 9.6). The average spatial variability in the whole brain was calculated to be 0.075 ± 0.050 mm. The highest degree of variability across subjects, up to as 0.3 mm, was seen in the olfactory bulb. Tissues around the *LV* and central part of the *cc* also tended to have large variability (up to 0.2 mm).

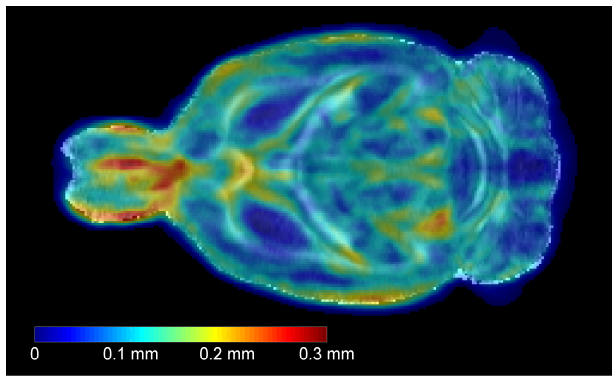


Figure 9.6: A single horizontal slice of the *in vivo* population-averaged FA template (in gray color scale) overlaid with the anatomical variability magnitude (AVM) map (shown in spectral color scale). The intensity levels in the AVM map denote distances in millimeters, and represent the average anatomical variability in the adult rat brain across the sample population. Blue regions indicate low anatomical variability, while red indicate high anatomical variability.

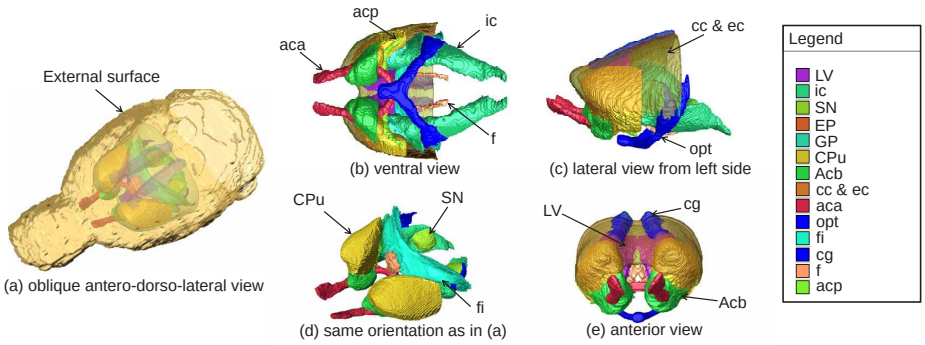


Figure 9.7: 3D rendering images of the delineated brain structures: the nucleus accumbus (Acb), caudate putamen complex (CPu), globus pallidus (GP), entopeduncular nucleus (EP), substantia nigra (SN), external capsule (ec), corpus callosum (cc), internal capsule (ic), fimbria of the hippocampus (fi), fornix (f), posterior (acp) and anterior part (aca) of anterior commisure, optic tract (opt), cingulum (cg), and lateral ventricle (LV) were manually segmented with ITK-SNAP and visualized using AMIRA software®.

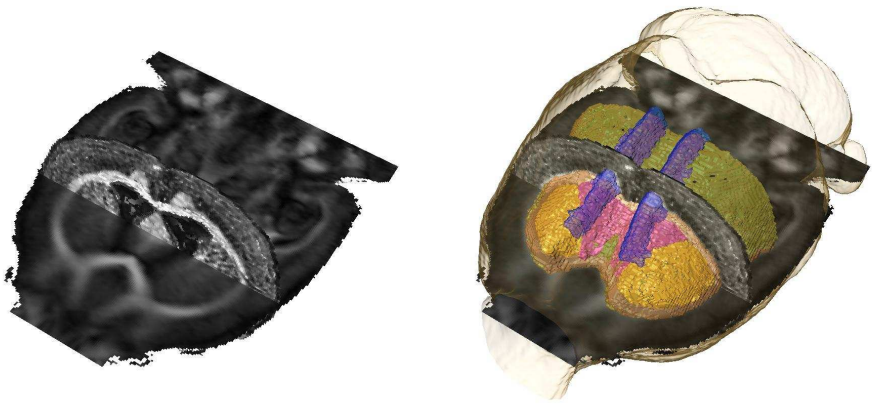
9.3.2 Manual delineation of ex vivo data

A selection of white matter and gray matter regions were manually segmented on basis of T_1-weighted (Fig. 9.2a) and a direction encoded (DEC) FA map that more effectively visualizes the directional information embedded in the primary eigenvector, i.e., the local fiber orientation (Fig. 9.2b). The boundaries of several structures that are difficult to distinguish in anatomical images are well differentiated in the colored DEC map. Two examples are visualized in Fig. 9.3. First, it is not possible to distinguish the cg and cc in the T_1 images, the DEC map however reveals that the fiber orientation of the cg, as running anterioposteriorly (blue), is almost perpendicular to the main direction of corpus callosum, which runs mediolateral (red). Second, the dorsal boundary of the ec is ambiguous in T_1-weighted images (arrowheads in (9.3(d)), but readily identified mediolaterally oriented diffusion orientations in DTI.

In Fig. 9.7, 3D rendering images of the delineated brain structures - Acb, CPu, GP, EP, SN, cc/ec, ic, fi, f, acp, aca, opt, cg, and LV - were visualized using AMIRA software®.

9.3.3 Correction of ex vivo data

A nonrigid coregistration algorithm was used to minimize the morphological differences between *in vivo* and *ex vivo* samples. In Fig. 9.8a, a horizontal and coronal section of the FA maps of the *in vivo* population based atlas and the *ex vivo* subject, respectively, are shown to indicate the accurate alignment of both datasets. In Fig. 9.8b, a 3D rendering of cc/ec, cg, and LV are superimposed onto Fig. 9.8a. The proper alignment of the anatomical labels, the *in vivo*, and *ex vivo* data is evident from Fig. 9.8. The corrected *ex vivo* atlas, i.e. the deformed *ex vivo* DTI and T_1 data and the corresponding parcellation map, can be obtained by contacting the corresponding author. Data are provided in Amira® and NifTi format.



(a) (b)

Figure 9.8: A horizontal and coronal section of the FA maps of the *in vivo* population based atlas and the *ex vivo* subject, respectively, are shown (a) to indicate the accurate alignment of both datasets after applying a nonrigid coregistration algorithm to minimize morphological differences between both samples. In (b), a 3D rendering of *cc*/*ec*, *cg*, and *LV* are superimposed onto (a); Coloring in accordance to Fig. 9.7.

9.4 Discussion

In this study, we have proposed an anatomically labeled DTI atlas of the average brain of the adult SD rat, the most widely used rat strain in laboratory animal research. Furthermore, the SD rat is so far the preferred strain to produce transgenic rat models of human neurodegenerative pathologies [e.g. Bugos et al., 2009, von Hörsten et al., 2003, Vermeiren et al., 2006, Liu et al., 2008].

The construction of brain atlases has been a topic of intense research for the last decades. Standard rat brain atlases consist of drawings based on histological studies of a Wistar and SD rat brain by Paxinos and Watson [1986, 2007] and Swanson [1992, 1998, 2004], respectively. The atlases describe the brain as a series of 2D sections with boundaries of areas and nuclei indicated, and names assigned to the delineated structures. Nevertheless, the use of such atlases for neuroimaging studies are hampered by the 2D format, resulting in discontinuous 3D structures after coarse alignment of the individual slices, as well as a lack of an analytical power for comparison of atlas and image data. While the diagrams presented in the atlases of Paxinos and Watson [1986, 2007] and Swanson [1992, 2004] are developed from study of several animals, they do not systematically incorporate anatomical variation across subjects within the given strain. Next, the lack of CSF pressure, and skull encasement constraints inherent to histological processing affect the brain, as a result of which morphological differences occur between the histology-derived atlases and *in vivo* data acquired with, e.g., MRI [Schwarz et al., 2006].

It has long been appreciated that MRI can advance the construction of brain atlases since it is a non-invasive, 3D and *in vivo* imaging technique, which enables accurate identification of a large number of anatomical structures. Therefore,

recently, some MRI templates of the rat brain and digital atlases of its major brain structures have been developed. First, Schweinhardt et al. [2003] constructed an MRI template that could be used to normalize subjects to the stereotaxic space defined by Paxinos and Watson [2007]. They created an MRI template based on high resolution T_2-weighted images of five female SD rats. A large set of anatomical landmarks were manually identified in each set of images and in the atlas. For each animal, an affine transformation that maps the set of landmarks of the corresponding image volume onto the atlas landmarks was estimated. These transformation matrices were used to resample the individual volumes into the stereotaxic space. Schwarz et al. [2006] constructed a stereotaxic T_2 template using 97 adult male SD rats. In addition, tissue classes were developed to guide delineation of the brain parenchyma from CSF. The MRI template was spatially normalized to the stereotaxic space of the Paxinos atlas by affine registration of the MRI brain tissue class map to the outline contour atlas images, corresponding to Paxinos' figures 4 - 78 [Paxinos and Watson, 2007]. Furthermore, Hess et al. [2005] aimed an automatic identification and structure assignment of activated voxel groups from functional MRI by using a labeled standard atlas. Therefore, they constructed a T_2 atlas from 54 SD rats using an affine coregistration technique to transform the individual subjects to the common reference template.

Our work distinguishes from the preceding studies concerning rat atlas construction since (I) a population-averaged brain atlas was constructed to obtain a template that represents the unbiased average anatomy. (II) During the construction, a nonrigid coregistration technique was used to avoid local misalignment inaccuracies, and partial volume averaging of anatomically distinct structures, due to intersubject morphological differences as shown in Fig. 9.4. (III) Accurate manual delineation of brain white matter structures was conducted on high resolution *ex vivo* diffusion weighted scans.

A high-resolution population-averaged MRI based atlas has a wide range of potential use, for example, as a reference space for coregistration of brain image data and for assigning anatomical labels to such data [Hjornevik et al., 2007]. Our DTI atlas is particularly useful tool for quantitative DTI group analyses, often based on a exploratory approach with voxel based statistical comparison or a hypothesis-driven manual ROI analysis. For voxels based comparisons our atlas provides an excellent spatial reference template for coregistration, while during ROI analyses, our atlas can be used to guide the manual delineation of anatomical structures. We have successfully conducted such analyses in studies on diffusion MRI changes in rats transgenic for Huntington disease [Antonsen et al., 2010]. To allow even more consistent and reproducible ROI delineations, one could use the DTI atlas to enable automated atlas-based delineation. For such purposes, the atlas is warped to the studied images by a spatial transformation, such that labels defined in the atlas are accurately projected onto the anatomically corresponding structures in the images under study. For ROI based comparisons of disease models, it is of importance that disease related changes are not averaged out by warping the data to an atlas. Therefore, we advise to apply the DTI atlas for standardized and automated ROI delineation by first warping brain volumes to the atlas, and subsequently use the inverse deformation field to map the atlas labels to the original data volume. An important source of inaccurate atlas-based delineation of ROIs includes registration

errors. Therefore, it is important to note that affine registrations often lead to insufficient correlation between manual and automated ROI approaches due to nonlinear intersubject anatomical or age dependent differences. Therefore, we advise to use the atlas in combination with a nonlinear coregistration algorithm. Although in-house coregistration software based on the viscous fluid model was used in the study, many other algorithm - often freely distributed - might be useful. Commonly used software packages are SPM and FSL in which the nonrigid deformations are defined by a linear combination of 3D discrete cosine transform, respectively, cubic B-spline basis functions [Ashburner and Friston, 1999, Smith et al., 2004]. A more thorough overview and evaluation of the broad range of nonrigid coregistration algorithms was given by Klein et al. [2009].

We would like to emphasize that the diffusion parameters that can be derived from the atlas are not ground truth values that can be used as a reference in future DTI studies. Although the diffusion of water molecules is a physical property of the tissue being measured, the estimated diffusion coefficients depend on scanner settings such as the b-value and, thus, the comparison between various DTI studies get hampered [e.g. Horsfield, 2001]. Veraart et al. [2011] demonstrated that a more accurate and b-value independent estimation of the diffusion parameters can be obtained with diffusion kurtosis imaging (DKI), a recently proposed higher order diffusion model [Jensen et al., 2005]. However, given a fixed acquisition time, an equally precise parameter estimation with DKI requires a decreased resolution, which might cause a lack of anatomical detail. Note that we preferred to put emphasis on the anatomical detail instead of the quantification accuracy in this study.

Future developments might include delineations of more anatomical structures and establishment of a spatial reference coordinate system based on internal anatomical landmarks, in line with the criteria used for defining the Waxholm space in the adult, male C57BL/6 mouse brain [Hawrylycz et al., 2009, 2011, Johnson et al., 2010]. This will further increase the application value of the atlas for integration and comparison of different data modalities.

Bibliography

M. Aggarwal, J. Zhang, M.I. Miller, R.L. Sidman, and S. Mori. Magn reson imaging and micro-computed tomography combined atlas of developing and adult mouse brains for stereotaxic surgery. *Neuroscience*, 162(4):1339 – 1350, 2009.

D.C. Alexander, C. Pierpaoli, P.J. Basser, and J.C. Gee. Spatial transformations of diffusion tensor magnetic resonance images. *IEEE Trans Med Imaging*, 20(11): 1131–1139, 2001.

D. Amaral and P. Lavenex. Hippocampal neuroanatomy. In Per Andersen, Richard Morris, David Amaral, Tim Bliss, and John O'Keefe, editors, *The Hippocampus*, chapter 3. Oxford University Press, New York, 2007.

B.T. Antonsen, Y. Jiang, Y. Veraart, J. andVan Dongen, H. Qu, Y. Urbach, K. Raber, H.P. Nguyen, J. Sijbers, S. von Hörsten, G.A. Johnson, and T.B.

Leergaard. Altered ex vivo diffusion mri in basal ganglia of transgenic huntington's disease rats. In 7^{th} *Forum of European Neuroscience Abstracts 0.46.2*, 2010.

J. Ashburner and K.J. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7:254–266, 1999.

P.J. Basser, J. Mattiello, and D. Le Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B*, 103(3):247–254, 1994.

O. Bugos, M. Bhide, and N. Zilka. Beyond the rat models of human neurodegenerative disorders. *Cell Mol Neurobiol*, 29(6-7):859–869, 2009. ISSN 1573-6830.

M.A. Cenci, I.Q. Whishaw, and T. Schallert. Animal models of neurological deficits: how relevant is the rat? *Nat Rev Neurosci*, 3(7):574–579, 2002. 574-579.

K.A. Coleman, G.E. Baker, and J. Mitrofanis. Topography of fibre organisation in the corticofugal pathways of rats. *J Comp Neurol*, 381(2):143–157, 1997.

J. Cozzi, A. Fraichard, and K. Thiam. Use of genetically modified rat models for translational medicine. *Drug Discov Today*, 13(11-12):488 – 494, 2008.

E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. A viscous fluid model for multimodal non-rigid image registration using mutual information. *Medical Image Analysis*, 7(4):565 – 575, 2003.

J. Dauguet, S. Peled, V. Berezovskii, T. Delzescaux, S.K. Warfield, R. Born, and C.F. Westin. Comparison of fiber tracts derived from in-vivo DTI tractography with 3D histological neural tract tracer reconstruction on a macaque brain. *NeuroImage*, 37(2):530–538, 2007.

C.R. Gerfen. Basal ganglia. In George Paxinos, editor, *The rat nervous system, third edition*. Elsevier Academic Press, San Diego, 2004.

M. Hawrylycz, J. Boline, A. Burger, T. Hashikawa, G.A. Johnson, M. Martone, L. Ng, J. Nissanov, L. Puelles, S. Ruffins, et al. The INCF digital atlasing program: Report on digital atlasing standards in the rodent brain. 2009.

M. Hawrylycz, R.A. Baldock, A. Burger, T. Hashikawa, G.A. Johnson, M. Martone, L. Ng, C. Lau, S.D. Larsen, J. Nissanov, L. Puelles, S. Ruffins, F. Verbeek, I. Zaslavsky, and J. Boline. Digital atlasing and standardization in the mouse brain. *PLoS Comput Biol*, 7(2):e1001065, 02 2011.

L. Heimer, F.F. Ebner, and W.J.H. Nauta. A note on the termination of commissural fibers in the neocortex. *Brain Res*, 5(2):171 – 177, 1967.

A. Hess, S. Kreitz, and K. Brune. *Functional atlas of the rat brain*, pages 73–77. Springer Berlin Heidelberg, 2005.

T. Hjornevik, T.B. Leergaard, D. Darine, O. Moldestad, A.M. Dale, F. Willoch, and J.G. Bjaalie. Three-dimensional atlas system for mouse and rat brain imaging data. *Frontiers in Neuroinformatics*, 1:4, 2007.

M.A. Horsfield. Using diffusion-weighted MRI in multicenter clinical trials for multiple sclerosis. *J Neurol Sci*, 186(Supplement 1):S51 – S54, 2001.

C.A. Ingham, S.H. Hood, P. Taggart, and G.W. Arbuthnott. Plasticity of synapses in the rat neostriatum after unilateral lesion of the nigrostriatal dopaminergic pathway. *J Neurosc*, 18(12):4732–4743, 1998.

M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825 – 841, 2002.

J.H. Jensen, J.A. Helpern, A. Ramani, H. Lu, and K. Kaczynski. Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magn reson imaging. *Magn Reson Med*, 53(6):1432–40, June 2005.

Y. Jiang and E.W. Hsu. Accelerating MR diffusion tensor imaging via filtered reduced-encoding projection-reconstruction. *Magn Reson Med*, 53(1):93–102, 2005. 93-102.

Y. Jiang, K. Pandya, and O. Smithies. Three-dimensional diffusion tensor microscopy of fixed mouse hearts. *Magn Reson Med*, 52(3):453–460, 2004. 453-460.

G.A. Johnson, G.P. Cofer, S.L. Gewalt, and L.W. Hedlund. Morphologic Phenotyping with MR Microscopy: The Visible Mouse. *Radiology*, 222(3):789, 2002.

G.A. Johnson, A. Ali-Sharief, A. Badea, J. Brandenburg, G. Cofer, B. Fubara, S. Gewalt, L.W. Hedlund, and L. Upchurch. High-throughput morphologic phenotyping of the mouse brain with magnetic resonance histology. *NeuroImage*, 37(1), 2007. 82-89.

G.A. Johnson, A. Badea, J. Brandenburg, G. Cofer, B. Fubara, S. Liu, and J. Nissanov. Waxholm Space: An image-based reference for coordinating mouse brain research. *NeuroImage*, 53(2):365–372, 2010.

D.K. Jones. Studying connections in the living human brain with diffusion MRI. *Cortex*, 44(8):936–952, 2008. ISSN 0010-9452.

J.A. Kaufman, E.T. Ahrens, D.H. Laidlaw, S. Zhang, and J.M. Allman. Anatomical analysis of an aye-aye brain (Daubentonia madagascariensis, Primates: Prosimii) combining histology, structural magnetic resonance imaging, and diffusion-tensor imaging. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology*, 287(1):1026–1037, 2005.

A. Klein, J. Andersson, B.A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G.E. Christensen, D.L. Collins, J. Gee, P. Hellier, J.H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R.P. Woods, J.J. Mann, and R.V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786 – 802, 2009. ISSN 1053-8119.

P. Kochunov, J.L. Lancaster, P. Thompson, R. Woods, J. Mazziotta, J. Hardies, and P. Fox. Regional spatial normalization: toward an optimal target. *J Comput Assist Tomogr*, 25(5):805–816, 2001. ISSN 0363-8715.

D. Le Bihan, J.-F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *J Magn Reson Imaging*, 13(4):534–546, 2001.

T.B. Leergaard, N.S. White, A. De Crespigny, I. Bolstad, H. D'Arceuil, J.G. Bjaalie, and A.M. Dale. Quantitative histological validation of diffusion MRI fiber orientation distributions in the rat brain. *PLoS ONE*, 5(1):e8595, 2010. doi: 10.1371/journal.pone.0008595.

C.P. Lin, W.Y.I. Tseng, H.C. Cheng, and J.H. Chen. Validation of diffusion tensor magnetic resonance axonal fiber imaging with registered manganese-enhanced optic tracts. *NeuroImage*, 14(5):1035–1047, 2001.

L. Liu, I.J. Orozco, E. Planel, Y. Wen, A. Bretteville, P. Krishnamurthy, L. Wang, M. Herman, H. Figueroa, W.H. Yu, Ottavio Arancio, and Karen Duff. A transgenic rat that develops alzheimer's disease-like amyloid pathology, deficits in synaptic plasticity and cognitive impairment. *Neurobiol Dis*, 31(1):46 – 57, 2008. ISSN 0969-9961.

F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging*, 16(2):187–198, 1997.

N.G. Papadakis, D. Xing, and C.L.H. Huang. A comparative study of acquisition schemes for diffusion tensor imaging using MRI. *J Magn Reson*, 137(1), 1999. 67-82.

G. Paxinos and C. Watson. *The Rat Brain in Stereotaxic Coordinates*. Sydney: Academic Press, 1986.

G. Paxinos and C. Watson. *The rat brain in stereotaxic coordinates*. Academic press, 2007.

M.F. Sargon, N. Mas, S. Senan, B. Ozdemir, H.H. Celik, and M. Cumhur. Quantitative analysis of myelinated axons of commissural fibers in the rat brain. *Anat Histol Embryol*, 32(3):141–144, 2003.

A.J. Schwarz, A. Danckaert, T. Reese, A. Gozzi, G. Paxinos, C. Watson, E.V. Merlo-Pich, and A. Bifone. A stereotaxic MRI template set for the rat brain with tissue class distribution maps and co-registered anatomical atlas: Application to pharmacological MRI. *NeuroImage*, 32(2):538–550, 2006.

P. Schweinhardt, P. Fransson, L. Olson, C. Spenger, and J.L. Andersson. A template for spatial normalisation of MR images of the rat brain. *J Neurosci Methods*, 129(2):105–113, 2003. ISSN 0165-0270.

S.M. Smith, M. Jenkinson, M.W. Woolrich, C.F. Beckmann, T.E.J. Behrens, H. Johansen-Berg, P.R. Bannister, M. De Luca, I. Drobnjak, D.E. Flitney, R.K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, and P.M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(Supplement 1):S208 – S219, 2004.

L.W. Swanson. *Brain maps: structure of the rat brain.* Elsevier Amsterdam, 1992.

L.W. Swanson. *Brain maps: structure of the rat brain: a laboratory guide with printed and electronic templates for data, models and schematics.* Elsevier Amsterdam, second edition, 1998.

L.W. Swanson. *Brain Maps: Structure of the Rat Brain. A Laboratory Guide with Printed and Electronic Templates for Data, Models and Schematics.* Elsevier,Amsterdam, third edition, 2004.

L.W. Swanson and W.M. Cowan. The connections of the septal region in the rat. *J Comp Neurol*, 186(4):621–655, 1979.

W.D. Van de Berg, A. Blokland, A.C. Cuello, C. Schmitz, W. Vreuls, H.W. Steinbusch, and C.E. Blanco. Perinatal asphyxia results in changes in presynaptic bouton number in striatum and cerebral cortex-a stereological and behavioral analysis. *J Chem Neuroanat*, 20(1):71–82, 2000.

W. Van Hecke, A. Leemans, E. D'Agostino, S. De Backer, E. Vandervliet, P. M. Parizel, and J. Sijbers. Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information. *IEEE Trans Med Imaging*, 26(11): 1598–1612, 2007.

W. Van Hecke, J. Sijbers, E. D'Agostino, F. Maes, S. De Backer, E. Vandervliet, P.M. Parizel, and A. Leemans. On the construction of an inter-subject diffusion tensor magnetic resonance atlas of the healthy human brain. *NeuroImage*, 43(1): 69–80, 2008.

J. Veraart, D.H.J. Poot, W. Van Hecke, I. Blockx, A. Van der Linden, M. Verhoye, and J. Sijbers. More accurate estimation of diffusion tensor parameters using diffusion Kurtosis imaging. *Magn Reson Med*, 65(1):138–45, 2011. ISSN 1522-2594.

C. Vermeiren, I. Hemptinne, N. Vanhoutte, S. Tilleux, J.-M. Maloteaux, and E. Hermans. Loss of metabotropic glutamate receptor-mediated regulation of glutamate transport in chemically activated astrocytes in a rat model of amyotrophic lateral sclerosis. *J Neurochem*, 96(3):719–731, 2006.

S. von Hörsten, I. Schmitt, H.P. Nguyen, C. Holzmann, T. Schmidt, T. Walther, M. Bader, R. Pabst, P. Kobbe, J. Krotova, D. Stiller, A. Kask, A. Vaarmann, S. Rathke-Hartlieb, J.B. Schulz, U. Grasshoff, I. Bauer, A.M.M. Vieira-Saecker, M. Paul, L. Jones, K.S. Lindenberg, B. Landwehrmeyer, A. Bauer, X.-J. Li, and O. Riess. Transgenic rat model of Huntington's disease. *Hum Mol Genet*, 12(6): 617–624, 2003.

P. Voorn, L.J. Vanderschuren, H.J. Groenewegen, T.W. Robbins, and C.M. Pennartz. Putting a spin on the dorsal-ventral divide of the striatum. *Trends Neurosci*, 27(8):468–474, 2004.

Q. Wang, D. Seghers, E. Dagostino, F. Maes, D. Vandermeulen, P. Suetens, and A. Hammers. Construction and validation of mean shape atlas templates for atlas-based brain image segmentation. *LNCS 3565*, pages 689–700, 2005.

L.E. Jr. White. Ipsilateral afferents to the hippocampal formation in the albino rat. *J Comp Neurol*, 113:1 – 14, 1959.

M. Woelche. Eine neue Methode der Markscheidenfarbung. *J Psychol Neurol*, 51: 199–202, 1942.

P.A. Yushkevich, J. Piven, H.C. Hazlett, R.G. Smith, S. Ho, J.C. Gee, and G. Gerig. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*, 31(3):1116–1128, 2006.

Conclusion

In this dissertation, we studied and improved the performance – both in terms of accuracy and precision – of diffusion MRI (dMRI) parameter estimators. To date, accurate estimation of diffusion model parameters is challenging due to the non-Gaussian distribution of magnitude MR data. Nonetheless, high accuracy can be achieved, often even with the class of linear least squares (LLS) estimators that is routinely used in DTI and DKI analyses. Note that the DTI and DKI models have in common that they can be structured into a linear regression form depending on the natural logarithm of the diffusion-weighted MR signals. However, the performance of the LLS estimators will depend on the SNR of the diffusion-weighted data. It is important to consider following properties:

1. If the measurement errors have expectation zero, then the LLS estimator is unbiased.

2. If the measurement errors have expectation zero and known variance, then the WLLS estimator for which the weight terms are the inverse of the variance of the respective measurement is the best linear unbiased estimator.

3. If the measurement errors are normally distributed with zero expectation and known variance, then the WLLS estimator for which the weight terms are the inverse of the variance of the respective measurement is the minimum variance unbiased estimator (MVUE).

First, in Chapter 5, we showed that property 1 only holds under some conditions. If the acquired diffusion-weighted MR data are Rice distributed and the SNR is at least two for all data samples, then the LLS estimators are unbiased. This property does not hold for the ordinary nonlinear least squares (NLS) estimator, which fit the diffusion models to the data in their native space. So, basically, the accuracy of ordinary least squares estimators improves by the log-transformation.
Second, to meet the conditions stated in property 2, the weights of the WLLS estimator should be the square of the noise-free diffusion weighted signals. Obviously, the noise-free diffusion-weighted signals are not known and, as such, the weight matrix needs to be estimated. In Chapter 5, we showed the importance of a well-considered choice of the weight matrix. Most importantly, correlations between the weights and the respective diffusion-weighted samples must be avoided. Otherwise, the improved precision comes with a drop in accuracy.
Third, if the SNR of diffusion-weighted samples is low to moderate, then the error term will not be normally distributed. This observation has some important limitations w.r.t. the precision of the WLLS estimator. Indeed, the WLLS estimator is generally not the MVUE. Moreover, the NLS estimator has typically an higher

precision. As such, choosing between NLS and WLLS is often not trivial considering that accuracy and precision are both important properties of an estimator. Theoretically, the maximum likelihood estimator (MLE) is a good alternative to least squares estimation, irrespective of SNR. This nonlinear estimator has optimal asymptotical properties regarding accuracy and precision, but it requires the analytical expression of the data PDF. However, there are two practical concerns regarding the use of MLE in the context of dMRI. First, the analytical expression of the PDF is based on the noise level. Although some methods suggested estimating the noise level as part of the model fit, we showed in Chapter 6 that the noise level is preferably estimated prior to model fitting because the accuracy and precision of diffusion model parameter estimators depend on the knowledge of this noise level. Nowadays, the estimation of the noise level is challenging due to the use of parallel imaging techniques. Indeed, the noise level has become spatially varying. An algorithm to estimate such 3D noise map is proposed in Chapter 7. Second, the necessity of data correction (e.g. motion and eddy current corrections) prior to model fitting causes the MLE's dependency on the data PDF to become a weakness because the altered data PDF can no longer be expressed analytically. A practical – potentially slightly less precise – alternative to the MLE, i.e. the conditional least squares estimator, is introduced in Chapter 6.

Finally, due to noise and imaging artifacts, it is often necessary to constrain the parameter estimation to guarantee biological and physical plausibility. In Chapter 8, we showed how the estimators, introduced/discussed in Chapters 5 and 6, can be constrained. Moreover, we showed that constraining the parameter estimation yields improved performance in terms of MSE in estimating the DKI model parameters.

List of Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
ACS	Autocalibration signal
AD	Axial diffusivity
ADC	Apparent diffusion coefficient
AK	Axial kurtosis
AKC	Apparent kurtosis coefficient
BLUE	Best linear unbiased estimator
CC	Corpus callosum
CHARMED	Composite hindered and restricted model of diffusion
CLS	Conditional least squares
CSF	Cerebrospinal fluid
DKI	Diffusion Kurtosis Imaging
DKT	Diffusion kurtosis tensor
dMRI	Diffusion (weighthed) MRI
DT	Diffusion tensor
DTI	Diffusion Tensor Imaging
DW	Diffusion-weighted
EPI	Echo planar imaging
FA	Fractional anisotropy
FOV	Field-of-view
FWHM	Full width half maximum
GM	Gray matter
GRAPPA	Generalized autocalibrating partially parallel acquisition
KA	Kurtosis anisotropy
LLS	Linear least squares
MAD	Median absolute deviation
MD	Mean diffusivity
MDT	Minimal deformation target
MK	Mean kurtosis
MLE	Maximum likelihood estimator
MRI	Mangetic resonance imaging
MSE	Mean squared error
mSENSE	Modified SENSE
MVUE	Minimum variance unbiased estimator
NEX	Number of experiments
NLS	Nonlinear least squares

NMR	Nuclear magnetic resonance
PDF	Probability distribution function
PGSE	Pulsed gradient spin echo
pMRI	Parallel MRI
PPD	Preservation principal direction
QP	Quadratic programming
RD	Radial diffusivity
RF	Radio frequency
RK	Radial kurtosis
ROI	Region of interest
SENSE	Sensitivity encoding
SNR	Signal-to-noise ratio
SPG	Short pulse gradient
SQP	Sequential quadratic programming
ss-EPI	Single-shot EPI
TE	Echo time
TR	Repetition time
WLLS	Weighted linear least squares
WM	White matter

List of Figures

Academic Overview

Journal papers

1. W. Van Hecke, A. Leemans, C. Sage, **J. Veraart**, J. Sijbers, S. Sunaert, and P.M. Parizel. The effect of template selection on diffusion tensor voxel based analysis results. *NeuroImage*, 55(2):566-573, 2010.
2. **J. Veraart**, D.H.J. Poot, W. Van Hecke, I. Blockx, A. Van der Linden, M. Verhoye, and J. Sijbers. More accurate estimation of diffusion tensor parameters using diffusion kurtosis imaging. *Magnetic Resonance in Medicine*, 65(1):138-145, 2011.
3. **J. Veraart**, W. Van Hecke, and J. Sijbers. Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a Rician noise model. *Magnetic Resonance in Medicine*, 66(3):678-686, 2011.
4. **J. Veraart**, T.B. Leergaard, B.T. Antonsen, W. Van Hecke, I. Blockx, B. Jeurissen, Y. Jiang, A. Van der Linden, G.A. Johnson, M. Verhoye, and J. Sijbers.Population-averaged diffusion tensor imaging atlas of the Sprague Dawley rat brain. *NeuroImage*, 58(4):975-983, 2011.
5. N. Van Camp, I. Blockx, L. Camón, N. de Vera, M. Verhoye, **J. Veraart**, W. Van Hecke, E. Martínez, G. Soria, J. Sijbers, A.M. Planas, and A. Van der Linden. A complementary diffusion tensor imaging DTI-histological study in a model of Huntington's disease. *Neurobiology of Aging*, 33(5):945-959, 2012.
6. S. Van Cauter, **J. Veraart**, J. Sijbers, R.R. Peeters, U. Himmelreich, F. De Keyzer, S.W. Van Gool, F. Van Calenbergh, S. De Vleeschouwer, W. Van Hecke, and S. Sunaert. Gliomas: diffusion kurtosis MR imaging in grading. *Radiology*, 263(2):492-501, 2012.
7. I. Blockx, M. Verhoye, J. Van Audekerke, I. Bergwerf, J.X. Kane, R. Delgado y Palacios, **J. Veraart**, B. Jeurissen, K. Raber, S. von Hörsten, P. Ponsaerts, J. Sijbers, T.B. Leergaard, A. Van der Linden. Identification and characterization of Huntington related pathology : an in vivo DKI imaging study *NeuroImage*, 63(2):653-662, 2012.
8. B.T. Antonsen, Y. Jiang, **J. Veraart**, H. Qu, H.P. Nguyen, J. Sijbers, S. von Hörsten, G.A. Johnson, and T.B. Leergaard. Altered diffusion tensor imaging measurements in aged transgenic Huntington disease rats. *Brain Structure and Function*, 218(3):767-778, 2012.
9. J. Rajan, **J. Veraart**, J. Van Audekerke, M. Verhoye, and J. Sijbers. Nonlocal maximum likelihood estimation method for denoising multiple-coil magnetic resonance images. *Magnetic Resonance Imaging*, 30(10), 1512-1518, 2012.

10. T. De Bondt, W. Van Hecke, **J. Veraart**, A. Leemans, J. Sijbers, S. Sunaert, Y. Jacquemyn, and P.M. Parizel. Does the use of hormonal contraceptives cause microstructural changes in cerebral white matter? Preliminary results of a DTI and tractography study. *European Radiology*, 23(1):57-64, 2013.
11. D.H.J. Poot, B. Jeurissen, Y. Bastiaensen, **J. Veraart**, W. Van Hecke, P.M. Parizel, and J. Sijbers. Super-resolution for multislice diffusion tensor imaging. *Magnetic Resonance in Medicine*, 69(1):103-113, 2013.
12. G. Vanhoutte, S. Pereson, R. Delgado y Palacios, P. Guns, B. Asselbergh, **J. Veraart**, J. Sijbers, M. Verhoye, C. Broeckhoven, and A. Van Der Linden. Diffusion kurtosis imaging to detect amyloidosis in an APP/PS1 mouse model for Alzheimer's disease. *Magnetic Resonance in Medicine*, 69(4):1115-1121, 2013.
13. S. Sekar, E. Jonckers, M. Verhoye, R. Willems, **J. Veraart**, J. Van Audekerke, J. Couto, M. Giugliano, K. Wuyts, S. Dedeurwaerdere, J. Sijbers, C. Mackie, L. Ver Donck, T. Steckler, and A. Van der Linden. Subchronic memantine induced concurrent functional disconnectivity and altered ultra-structural tissue integrity in the rodent brain: revealed by multimodal MRI. *Psychopharmacology*, 227(3), 479-491, 2013.
14. **J. Veraart**, J. Rajan, R. Peeters, A. Leemans, S. Sunaert, and J. Sijbers. Comprehensive framework for accurate diffusion MRI parameter estimation. *Magnetic Resonance in Medicine*, 2013, in press.
15. **J. Veraart**, J. Sijbers, S. Sunaert, A. Leemans, and B. Jeurissen. Weighted linear least squares estimation of diffusion MRI parameters: strenghts, limitations and pitfalls. *NeuroImage*, 81:335-346, 2013.

Conference papers

1. **J. Veraart**, W. Van Hecke, I. Blockx, A. Van Der Linden, M. Verhoye, and J. Sijbers. Non-rigid coregistration of diffusion kurtosis data. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, 392-395, 2010.
2. T. Dhollander, **J. Veraart**, W. Van Hecke, F. Maes, S. Sunaert, J. Sijbers, and P. Suetens. Feasibility and advantages of diffusion weighted imaging atlas construction in Q-space. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 166-173, 2011.

Conference abstracts (First author only)

1. **J. Veraart**, I. Blockx, W. Van Hecke, M. Verhoye, A. Van der Linden and J. Sijbers, Improved non rigid coregistration of diffusion kurtosis images by incorporating diffusion kurtosis tensor information, *Proceedings of the European Society for Magnetic Resonance in Medicine and Biology*, page 36, Antalya, Turkey, 2009
2. **J. Veraart**, B. T. Antonsen, I. Blockx, W. Van Hecke, Y. Jiang, G. A. Johnson A. Van Der Linden T. B. Leergaard, M. Verhoye and J. Sijbers.

Construction of a population based diffusion tensor image atlas of the Sprague Dawley rat brain. *In Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 18, page 3145, Stockholm, Sweden, 2010.

3. **Jelle Veraart**, W. Van Hecke, D. H.J. Poot, I. Blockx, A. Van der Linden, M. Verhoye and J. Sijbers. A more accurate and b-value independent estimation of diffusion parameters using Diffusion Kurtosis Imaging. *In Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 18, page 1687, Stockholm, Sweden, 2010.
4. **Jelle Veraart**, W. Van Hecke, D. H.J. Poot, and J. Sijbers. Constrained maximum likelihood estimator for more accurate diffusion kurtosis tensor estimates. *In Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 19, page 3947, Montreal, Canada, 2011.
5. **J. Veraart** and J. Sijbers, The bias-corrected least squares estimator: a fast and practical way to remove the Rician bias in diffusion MRI, *Proceedings of the European Society for Magnetic Resonance in Medicine and Biology*, page 498, Leipzig, Germany, 2011
6. **Jelle Veraart**, S. Van Cauter, S. Sunaert, and J. Sijbers. Conditional Least Squares Estimation of Diffusion MRI Parameters, *In Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 20, page 985, Melbourne, Australia, 2012
7. **Jelle Veraart** and J. Sijbers.The Rician Bias in Diffusion MRI: A Technical Overview, *In Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 20, page 3572, Melbourne, Australia, 2012
8. **J. Veraart**, J. Rajan, R.R. Peeters, S. Sunaert and J. Sijbers, Diffusion MRI: Estimation of spatially variable Rician noise, *Proceedings of the European Society for Magnetic Resonance in Medicine and Biology*, page 266, Lisbon, Portugal, 2012.
9. **J. Veraart**, J. Sijbers, A. Leemans, and B. Jeurissen, Ill-chosen weights weigh on the accuracy of the weighted linear least squares estimator in diffusion-weighted MRI, *Proceedings of the ISMRM Benelux Chapter*, volume 5, page 4-6, Rotterdam, The Netherlands, 2013
10. **Jelle Veraart** J. Rajan, R. Peeters , A. Leemans, S. Sunaert and J. Sijbers. Estimation of spatially variable Rician noise map in diffusion MRI, *In Proceedings of the International Society for Magnetic Resonance in Medicine*, volume 21, page 2081, Salt Lake City (UT), USA, 2013

Moderator

1. Scientific session: *Methodologic improvements in dMRI*, European Society for Magnetic Resonance in Medicine and Biology (ESMRMb), Lisbon, Portugal, 2012.

Invited talks at research groups

1. *More accurate estimation of diffusion MR parameters*, Center for Biomedical Imaging, Dept of Radiology, NYU Langone Medical Center, New York, USA, September 24, 2013

Relevant courses and workshops

1. *Diffusion: from Basic Physics to Exploration of Microscopic Structure*, Lectures on MR organized by the European Society for Magnetic Resonance in Medicine and Biology, Oxford, United Kingdom, October 14-16, 2010
2. *Diffusion as a Probe of Neural Tissue Microstructure*, organized by the International Society for Magnetic Resonance in Medicine, Podstrana, Croatia, October 14-18, 2013
3. Weekend Educational Program organized by the International Society for Magnetic Resonance in Medicine, several locations, 2010-2013

Awards

1. *Educational Stipend Award* for the submitted work, entitled: Constrained maximum likelihood estimator for more accurate diffusion kurtosis tensor estimates, at the 19th ISMRM meeting – Montreal, Canada, 2011.
2. *Educational Stipend Award* for the submitted work, entitled: Conditional Least Squares Estimation of Diffusion MRI Parameters, at the 20th ISMRM meeting – Melbourne, Australia, 2012.

Journal cover images

1. NeuroImage, Vol. 58(4), 2011
2. NeuroImage, Vol. 81, 2013

Reviewer for the following journals

1. American Journal of Neuroradiology
2. Human Brain Mapping
3. IEEE Transaction on Medical Imaging
4. Journal of Magnetic Resonance Imaging
5. Journal of Neuroscience Methods
6. Measurement Science and Technology
7. NeuroImage
8. Physics in Medicine and Biology

Reviewer for the following conferences

1. ISMRM Benelux chapter