Universiteit
Antwerpen

Faculteit Wetenschappen
Departement Fysica

Advances in the reconstruction and statistical processing of Magnetic Resonance images

---

Verbeteringen in de reconstructie en statistische verwerking van Magnetische Resonantie beelden

Proefschrift voorgelegd tot het behalen van de graad van

Doctor in de Wetenschappen

aan de Universiteit Antwerpen te verdedigen door

Ir. Dirk H. J. POOT

Promotoren:
Prof. Dr. Jan Sijbers
Dr. Ir. Arnold J. den Dekker

Antwerpen, juni 2010

Doctoral committee

Chairman
Prof. Dr. Jan Naudts

Members
Prof. Dr. Jan Sijbers
Dr. Ir. Arnold J. den Dekker
Prof. Dr. Paul Scheunders

External members
Prof. Dr. Wiero Niessen
Prof. Dr. Eric Achten

Contact information

✉ Dirk H. J. Poot
Biomedical Imaging Group Rotterdam
Erasmus Medical Center
Dr. Molewaterplein 50/60, Ee2112
3015 GE Rotterdam, the Netherlands
P.O. Box 2040, 3000 CA Rotterdam, the Netherlands

☎ +31 (0) 1070 43049

+31 (0) 6 1342 9595

+31 (0) 1070 44722

dirk.poot@ua.ac.be

d.poot@erasmusmc.nl

http://visielab.ua.ac.be/staff/poot
http://www.bigr.nl/people/DirkPoot

Published by: Uitgeverij BOXPress, Oisterwijk
Printed by: Proefschriftmaken.nl || Printyourthesis.com

Dankwoord

Deze thesis beschrijft het onderzoek dat ik heb uitgevoerd als onderdeel van mijn doctoraatsonderzoek aan het Visielab van de Universiteit Antwerpen. Dat onderzoek zou ik niet hebben kunnen doen zonder de hulp en steun van vrienden, familie en collega's. Als eerste wil ik daarom ook mijn promotor, Prof. Jan Sijbers, en co-promotor, Dr. Ir. Arjan den Dekker, hartelijk danken. Jan, hartelijk dank voor al de ondersteuning aan en, af en toe, het bijsturen van mijn onderzoek. Dank voor de gegeven vrijheid en het vertrouwen bij het invullen van het onderzoek. Ook wil ik u danken voor al het commentaar op de verschillende revisies van mijn papers en ook op deze thesis. Uw commentaar was altijd uitgebreid en het heeft sterk bijgedragen aan de kwaliteit en leesbaarheid van het geschrevene. Arjan, al weer heel wat jaren geleden kwam ik bij de groep "Model gebaseerd meten en regelen" voor mijn afstudeerproject. Hierbij was je mijn afstudeerbegeleider. Enige tijd na mijn afstuderen heb je me gewezen op de positie waarvoor Jan iemand zocht. Hartelijk dank daarvoor, want het was een goede en leerzame tijd die ik in Antwerpen heb gehad. Ook dank voor de begeleiding die je steeds hebt gegeven, met name tijdens de 'maandag' waarop ik in Delft was.

Naast mijn promotor en co-promotor zijn er nog vele mensen die mij tijdens mijn doctoraat in Antwerpen hebben geholpen en gestimuleerd. Als eerste wil ik al mijn collega's van het Visielab bedanken. Wim van Hecke, altijd kritisch, positief meedenkend. Hartelijk dank voor je bijdragen en open instelling. Ben Jeurissen, CSD is jouw specialiteit, maar ook bij andere dingen wilde je altijd helpen. Jelle Veraart, jij bent nu de DKI specialist op het Visielab. Succes met het nog verder promoten ervan en dank voor je collegialiteit. Arno Duijster, hartelijk dank voor de vriendelijkheid en collegialiteit. We zijn ongeveer tegelijk vanuit Delft naar Antwerpen gegaan. Iedere week hebben we samen gesquashed, om de conditie nog een beetje op peil te houden. Wouter Pintjens, helaas heb je je doctoraat niet kunnen afmaken, maar ik wil je hartelijk bedanken voor je werk en hoe je was. Ook in het bijzonder bedankt voor het opnemen van de speciaal geprogrammeerde MR opnamen die je hebt gemaakt. Alain Smolders, hartelijk dank voor de introductie in fMRI en de bijbehorende datasets. Prof. Paul Scheunders, bedankt voor de verschillende discussies en dat ik Wiskundige methoden voor de fysica mocht begeleiden. Het was nuttig,

ook voor mijn eigen onderzoek. Natuurlijk wil ik ook de andere collega's van het Visielab hartelijk bedanken voor de fijne sfeer die er altijd was.

Tijdens mijn doctoraat heb ik samengewerkt met mensen van het Bio Imaging lab. Met name Prof. Marleen Verhoye, Ines Blockx en Johan Van Audekerke. Jullie hebben de beelden van ratten en muizen opgenomen, hartelijk dank daarvoor en natuurlijk ook voor de goed discussies over mijn onderzoek. Verder wil ik ook Vincent Van Meir bedanken voor de opname van de multi-slice MR beelden, wat de aanleiding was voor een heel hoofdstuk in deze thesis.

Verder wil ik de mensen op de radiologie afdeling van het Universitair Ziekenhuis Antwerpen, met name Prof. Paul Parizel bedanken voor de opnamen van de humane MR beelden. Ook ben ik de afdeling radiologie van het Universitair Ziekenhuis Gent, en dan met name Prof. Rik Achten, mijn dank verschuldigd voor het opnemen van de verschillende DKI beelden. Wat u graag deed, ondanks de altijd grote drukte op de Radiologie afdeling.

Mijn ouders, Sjaak en Els wil ik bedanken voor het telkens stimuleren om te onderzoeken en natuurlijk ook voor hun liefde, steun, vertrouwen en interesse in mijn onderzoek. Ook wil ik mijn broer Bram en zussen Annelies en Marianne bedanken voor de gezelligheid en afleiding die af en toe ook nodig is.

Als allerlaatste, wil ik jou het allermeest bedanken, mijn lieve Teunitha. We hebben elkaar voor het eerst ontmoet toen ik al in Antwerpen bezig was. Sindsdien kon mijn doctoraat in Antwerpen voor jou niet snel genoeg zijn afgerond. Hartelijk dank voor de vreugde en steun die je me altijd geeft.

Zaltbommel, juni 2010,
Dirk Poot

Samenvatting

Medische Magnetische Resonantie beeldvorming is een belangrijk onderzoeksveld, in het bijzonder voor het bestuderen van het (menselijk) brein. Deze thesis bevat verschillende bijdragen aan dit onderzoeksveld. Het bevat verbeteringen van Magnetische Resonantie (MR) beelden alsook verbeterde statistische analyse van MR beelden die gevoelig zijn voor specifieke weefsel eigenschappen, zoals de diffusie van water of het zuurstofniveau van het bloed. Deze nieuwe methoden zijn er op gericht de gevoeligheid van de analyse van MR beelden te optimaliseren.

Hoofdstuk 1 bevat een introductie van MR beeldvorming, een beschrijving van enkele geconstateerde problemen bij MR beeldvorming en een opsomming van de bijdragen die geleverd zijn aan de oplossingen voor deze problemen.

Hoofdstuk 2 bevat een beschrijving van de verschillende methoden die op verschillende plekken in deze thesis zijn gebruikt.

Hoofdstuk 3 beschrijft een methode om de **inhomogeniteiten van het hoofdmagneetveld van de MR scanner te bepalen**. Deze inhomogeniteiten zijn vaak een hoofdoorzaak van verstoringen die optreden bij het opnemen van hele MR beelden na één echo, zogenaamde Echo Planar Imaging (EPI) opnamen. EPI is een heel snelle opnamemethode die, helaas, erg gevoelig is voor veldinhomogeniteiten. Echter, als de inhomogeniteiten van het hoofdmagneetveld bekend zijn, dan is het mogelijk om de verstoringen substantieel te reduceren. De nieuwe methode bepaalt de inhomogeniteiten van het magneetveld met behulp van een niet lineaire kleinste kwadraten schatter (NLLS) uit de complexwaardige gegevens die opgenomen werden met een aangepaste EPI opnamesequentie. Deze NLLS methode bleek veel minder gevoelig voor de meetruis dan een eerder gepresenteerde schatter voor de inhomogeniteiten van het magneetveld, welke was gebaseerd op de correlaties van de meetgegevens.

Hoofdstuk 4 beschrijft een nieuwe methode om **3D MR beelden met een hoge resolutie te reconstrueren** uit een serie MR beelden, ieder bestaand uit een aantal snedes. In zulke MR beelden is de dikte van de snedes doorgaans substantieel groter dan de resolutie in iedere snede. De ontwikkelde methode combineert meerdere beelden, ieder met een andere oriëntatie van de snedes, tot een hoog resolutie beeld van het gescande object. Binnen deze methode wordt een stelsel van lineare vergelijkingen opgesteld waarmee de opname

van MR beelden wordt gemodelleerd als een matrix-vector vermenigvuldiging. Vervolgens wordt het hoog resolutie MR beeld bepaald door dit -enorm grote- lineaire stelsel op te lossen met behulp van de 'geconjugeerde gradiënten' (conjugated gradient) methode. Om de matrix-vector vermenigvuldigingen op efficiënte wijze uit te voeren worden deze herschreven als combinatie van een filter operatie en een affiene transformatie van afbeeldingen. Onder enkele voorwaarden kan met behulp van deze nieuwe recontructie methode de opname duur worden verkort door meerdere MR beelden bestaande uit snedes op te nemen in plaats van directe 3D opnamen.

Hoofdstuk 5 beschrijft een methode om **automatisch het ruisniveau van MR beelden te bepalen**. Dit is relevant voor statistische analyse van MR beelden, aangezien het ruisniveau bekend moet zijn om onderscheid te kunnen maken tussen relevant signaal en ruis. De ontwikkelde methode is gebaseerd op het feit dat MR beelden meestal voor een relatief groot gedeelte uit achtergrond bestaan. In deze achtergrond is niet het object, maar alleen lucht aanwezig. Verder wordt gebruikt dat bij de meeste MR opnamen de ruis homogeen is in het beeld. Dat wil zeggen dat het ruisniveau in de achtergrond gelijk is aan het ruisniveau waarmee het object is verstoord. De intensiteit van de voxels uit de achtergrond, welke dus alleen ruis bevatten, is verdeeld volgens een Rayleigh verdeling. De methode in Hoofdstuk 5 gebruikt een maximale waarschijnlijkheid (maximum likelihood, ML)-schatter om de standaarddeviatie van de Rayleigh verdeelde achtergrond mode van histogram van een MR beeld te schatten. Het aantal elementen van het histogram waarmee deze schatter werkt wordt automatisch geselecteerd met behulp van een methode die probeert om een balans te zoeken tussen de variantie en onzuiverheid van de schatter. Simulatie experimenten laten zien dat de gemiddelde kwadratische fout van het geschatte ruisniveau is gereduceerd ten opzichte van meerdere eerder gepresenteerde methoden om het ruisniveau te bepalen.

Hoofdstuk 6 beschrijft op **waarschijnlijkheid gebaseerde statistische activatie detectie methoden voor functionele MRI**. Met behulp van functionele MRI kan de taakgerelateerde hersenactiviteit worden gedetecteerd uit een tijdreeks van MR afbeeldingen van de hersenen. Door fysiologische processen is de ruis die in deze tijdreeks van MR beelden zit gecorreleerd (gekleurd i.p.v. wit). Zoals in dit hoofdstuk wordt gedemonstreerd moeten de correlaties van de ruis nauwkeurig worden gemodelleerd om een juist niveau van foutief actief gedetecteerde voxels te kunnen specificeren. Verder toont dit hoofdstuk ook aan dat, zelfs als de ruis correct wordt gemodelleerd, het niveau van foutief actief gedetecteerde voxels af kan wijken van de gespecificeerde waarde; in het bijzonder voor datasets met een beperkt aantal elementen in de tijdreeks. Aangezien in de praktijk het aantal elementen in de tijdreeks altijd beperkt is, geven de verschillende activatie detectie methoden andere antwoorden bij reële datasets, terwijl ze asymptotisch identiek zijn. In Hoofdstuk 6 wordt de kwaliteit van de activatie detectie methoden onderzocht als functie van de foutief actief gedetecteerde voxels en de gevoeligheid van de detectie van

hersenactiviteit. Er wordt aangetoond dat de op waarschijnlijkheid gebaseerde testen een kleine verbetering geven ten opzichte van de testen met het algemeen lineaire model (GLM) , zelfs als bij deze testen gebruik wordt gemaakt van gedecorreleerde metingen.

Hoofdstuk 7 beschrijft een methode om de **opnameinstellingen voor diffusie kurtosis beelden (DKI) te optimaliseren**. Het DKI model is een recent ontwikkelde uitbreiding van het al bekende diffusie tensor model (DTI-model) dat Gaussische diffusie van watermoleculen veronderstelt. Door restricties van celwanden en andere structuren is de diffusie echter niet volledig Gaussisch. Daarom wordt binnen het DKI model de beschrijving van de verdeling van diffunderende watermoleculen uitgebreid met de kurtosis van deze verdeling. Door deze uitbreiding moeten diffusie gewogen beelden (DWI) met verschillende intensiteiten van diffusie weging (b-waarden) worden opgenomen voor een DKI-experiment. Door de b-waarden en de richtingen van de diffusie wegingsgradiënten te optimaliseren kan de precisie van de uit de MR beelden geschatte parameters van het DKI model worden gemaximaliseerd, voor een vooraf gegeven aantal DWI. In Hoofdstuk 7 wordt de precisie gekwantificeerd met behulp van de Cramér Rao ondergrens (CRLB). Dit is een theoretische ondergrens voor de variantie van parameters als ze worden geschat met een zuivere schatter. Met behulp van simulatie-experimenten wordt aangetoond dat de variantie van de DKI parameters, als ze worden geschat met de maximale waarschijnlijkheidsschatter (ML), deze CRLB (bijna) bereikt. Dit geldt zelfs voor de datasets met een beperkt aantal opnamen. Daarom kunnen de diffusie wegingsinstellingen betrouwbaar worden geoptimaliseerd door het minimaliseren van de CRLB van de DKI parameters. Opname instellingen die praktisch optimaal zijn worden gevonden met behulp van een gesimuleerde annealing (gesimuleerd uitgloeien) optimalisatie techniek. Wanneer deze geoptimaliseerde instellingen worden vergeleken met de meer traditionele instellingen blijkt dat de CRLB wordt gehalveerd door de instellingen te optimaliseren.

Summary

Medical Magnetic Resonance image processing is an important research topic, especially for studies of the (human) brain. This thesis contains several contributions to this field of research, both for the improvement of Magnetic Resonance (MR) images, as well as for the statistical analysis of MR images that are sensitive to specific tissue properties, such as water diffusion and blood oxygenation. These new methods aim to optimize the sensitivity of the analysis of MR images.

Chapter 1 contains a brief introduction into MR imaging, a description of problems that arise in MR imaging, and a description of the contributions that are made to solve these problems.

Chapter 2 introduces existing methods that are used throughout this work.

Chapter 3 describes a method to **measure the inhomogeneities of the main magnetic field of the MR scanner**. These inhomogeneities are often a main cause of distortions for the Echo Planar Imaging (EPI) method, which is a very fast acquisition method that, unfortunately, is sensitive to these field inhomogeneities. However, when the main magnetic field inhomogeneities are known, it is possible to significantly reduce these distortions. The new method estimates the field inhomogeneities with a non linear least squares (NLLS) estimator from the complex reference data that was acquired by a modified EPI sequence. This NLLS method proved to be much more robust against noise than a previously presented field inhomogeneity estimator that was based on correlations in the reference data.

Chapter 4 describes a new method to **reconstruct high resolution MR images** from a series of multi slice MR images with anisotropic resolution. In these multi slice images, the slice thickness is substantially larger than the resolution in the acquired slice. The method combines several multi slice images, each with a different slice orientation, to reconstruct a single high resolution image of the object. The reconstruction method first models the acquisition of the MR images by a linear system of equations. Next, the high resolution image is reconstructed by solving this -very large- linear system with the conjugated gradient method. The matrix vector multiplications present in the conjugated gradient method are reformulated as a combination of a filter operation and an affine transformation of images to obtain an efficient reconstruction method

for the high resolution image. With this reconstruction method, under certain circumstances, the acquisition duration can be reduced by acquiring several multi slice images of a 3D volume, instead of a true 3D acquisition.

Chapter 5 describes a new method to **automatically estimate the noise level** of magnitude MR images. This is relevant for statistical analysis of MR images, as the noise level needs to be known in order to be able to discriminate between signal and noise. The method that has been developed is based on the fact that MR images often contain a relatively large background area, i.e. an area in which the object is not present. In most acquisitions, the noise is homogeneous throughout the MR image. Therefore, the noise level that is present in the background area is equal to the noise level by which the intensities of the recorded object are corrupted. The magnitude of the voxels in the background area of MR images, which only contain noise, are known to be Rayleigh distributed. The method in Chapter 5 uses a Maximum Likelihood estimator to fit the Rayleigh distribution to the background mode of the histogram of an MR image. The number of histogram bins with which the Maximum Likelihood estimate of the noise variance is computed, is selected automatically by a method that tries to balance bias and variance of the estimator. Simulation experiments show that the root mean square error of the estimated noise level is reduced compared to previously presented methods.

Chapter 6 describes **likelihood based statistical activation tests for functional MRI**. With functional MRI, task related brain activation can be detected from a time series of brain images. However, due to physiological processes, the noise in the time series of MR images is often colored, instead of white. As is demonstrated in this chapter, the coloring of the noise should be appropriately modeled in order to correctly specify a false positive level. However, Chapter 6 also demonstrates that, even when the noise is appropriately modeled, the false positive level of the statistical activation tests might deviate from the intended value; especially for datasets with a limited number of time points. As in practice the number of time points is limited in all datasets, the activation tests, which are asymptotically equal, are not equal for real datasets. In Chapter 6, the performance of these different activation tests, for finite data length and with colored noise, is evaluated in terms of the false positive level and the sensitivity by which activation will be detected. It is shown that the likelihood based tests provide a slight improvement over the GLM test with pre-whitened data.

Chapter 7 describes a method to **optimize the acquisition of diffusion kurtosis images (DKI)**. The DKI model is a recently developed extension to the better known Diffusion Tensor imaging (DTI) model, which assumes Gaussian diffusion of water molecules. However, due to cell boundaries and other restrictions, the diffusion inside the brain is not purely Gaussian. In DKI, the model of the distribution of diffusing water molecules is extended to include the kurtosis of this distribution. Due to this extension, Diffusion Weighted images (DWIs) with different amounts of diffusion weighting (b-values) have to be ac-

quired in DKI experiments. By optimizing the b-values and diffusion weighting gradient directions, the precision of the DKI parameters can be maximized for any given number of DWI. In Chapter 7, the precision of the DKI parameters is quantified by the Cramér Rao lower bound (CRLB), which is a lower bound on the variance of these parameters when they are estimated by an unbiased estimator. With simulation experiments it is shown that the variance of DKI parameters estimated by the Maximum Likelihood estimator (almost) reaches the CRLB, even for the finite length datasets that are considered. Therefore, the diffusion weighting acquisition settings are optimized by minimizing the CRLB of the DKI parameters. Acquisition settings that are close to optimal are found by a simulated annealing optimization technique. A comparison of the resulting optimized settings with more traditional settings shows that the CRLB is halved by optimizing the settings.

Contents

Chapter 1

Introduction

Currently, the brain, and especially the human brain, is the focus of many studies. The aim of these studies is to increase the insights in the very complex anatomy and functionality of both the healthy and diseased brain. Until recently, the only techniques that were available to obtain information about the neural architecture and functionality of the human brain were histological post-mortem studies of the human brain or invasive studies on animals [1, 2]. However, recent advances in the Magnetic Resonance (MR) scanner hardware, as well as improvements in the analysis of the MR images, allowed advanced studies of the living human brain [3–5].

The aim of the work presented in this thesis is to improve and develop statistical methods for the acquisition and analysis of MR images in specific contexts. These methods focus on brain images, although some of the new methods can also be applied outside the area of brain imaging. As brain imaging is still a very broad topic, two applications of brain imaging were selected: Functional Magnetic Resonance Imaging (fMRI), which detects the active areas of the brain from a series of MR images, and Diffusion Weighted Imaging (DWI), with which the diffusion of water inside the brain is studied. In order to support the research for these applications, several other methods were developed. Especially, the distortions that are present in MR images were investigated, both spatial distortions as well as the noise with which the MR images are corrupted.

In this chapter, Magnetic Resonance imaging and Neuro Imaging are introduced first. Next, the problems that were identified and which are studied in this thesis are described. Finally, the main contributions of this thesis are

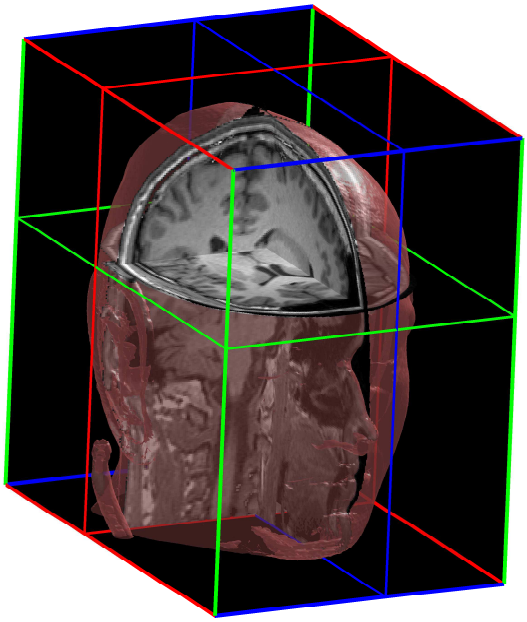


Figure 1.1: *T1 weighted MR image in sagittal (surrounded by the blue rectangle), coronal (red rectangle), and axial (green rectangle) views, combined with a semi transparent view of the surface of the head.*

summarized and the structure of this thesis is explained.

1.1 Magnetic Resonance Imaging

The brief introduction to Magnetic Resonance Imaging (MRI) that is given in this section should not be considered as replacement for the many good introductions to MRI which have previously been written, such as [6], [7].

The term 'Magnetic Resonance Imaging' consists of 3 parts:

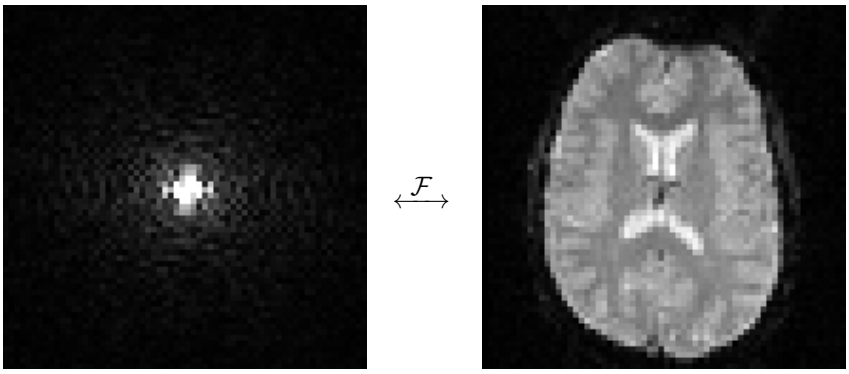
Imaging The aim of MRI is to obtain images of 'objects'. These 'objects' can really be objects, i.e. pieces of material, but in this context also a human (patient) or a (small) animal. The images that are recorded by an MR machine are not from the surface of an object, like a normal photograph, but also of the inside. The whole object, or a part of it, is scanned into a 3 dimensional (3D) image. Such 3D image consists of many volume elements, which are called voxels. See Fig. 1.1 for an example of a 3D MR image of the head of a human volunteer.

Magnetic The MR images are obtained by an MR scanner, which contains a large magnet. Currently, the magnetic magnetic field strength, which probably is the most important specification of a MR machine, is 1.5 Tesla (T) up to 7 T. This magnetic field is generated by a super conducting magnet, since superconducting magnets are the most efficient at the field strengths that are required. As a comparison, the magnetic field strength of the earth is 30 μT to 60 μT, so approximately 100 000 times weaker than the field strength of an MR magnet. A typical refrigerator magnet is around 70 mT, so still 100 times weaker.

Resonance The large strength of the magnet in the MR scanner causes the spins of nuclei of atoms with a net magnetic moment to align with the main magnetic field. In medical MRI, usually only the nucleus of hydrogen atoms, i.e. protons, are studied. The images are created by exciting the nuclear spin with radio waves of a specific frequency: the resonance frequency of the imaged nucleus. This resonance frequency depends on the magnetic field strength, is different for each type of nucleus, and is moderately sensitive to the environment of the atom, i.e. in what kind of molecule and tissue it is located. After excitation by the radio waves, the object starts to emit radio waves. These are recorded by radio receiver coils that are placed around the object. For the reconstruction of images, spatial information needs to be encoded into the recorded radio waves. This encoding is performed with spatially varying magnetic fields, i.e. magnetic gradients, which cause the resonance frequency to depend on the location. These magnetic gradients are generated by the gradient coils that are located inside the main MR magnet. By applying specific sequences of magnetic gradients, the (demodulated) radio waves can record the k-space of an image. This k-space is the domain of spatial frequencies and is related to the object space via the Fourier transform, see Fig. 1.2.

There are many properties of the object that can influence the images recorded by the MR scanner, such as proton density, tissue type, blood oxygenation, and water diffusion. The influence that these object properties have on the images depends on the specific sequence of radio waves and magnetic gradients, i.e. the type of acquisition, that is used to record the MR images. Therefore, many specialized acquisition types have been developed to record images. Each of these methods encodes specific information about the object under study in the image contrast.

Of the different methods developed in this thesis, several are valid for almost all MR images, regardless of the contrast by which they are recorded. These methods are presented in the Chapters 3, 4, and 5. In the Chapters 6 and 7, MR images with specific contrasts are studied.



(a) k-space magnitude (b) magnitude in image space

Figure 1.2: *One slice of a recorded MR image, in 2D k-space and 2D image space. These two images are related through the Fourier transform.*

1.2 Neuro imaging

The brain is a very complex organ that is still only partially understood. The uniquely human cognitive capabilities are attributed to the brain. Many diseases cause, or are caused by, disrupted processes in the brain. Therefore, a lot of studies focus on the detection and characterization of both normal and abnormal processes in the brain. The brain of animals, healthy human volunteers, and specific patient groups is studied in order to increase the insight in the very complex anatomy and functionality of the brain, and how the normal functionality is compromised in certain diseases. For these brain studies, MR imaging is extremely important as it provides, in contrast to almost every other method, the opportunity to study intact living brain tissue. The brain is not only easily damaged by surgery, it is also hard to obtain significant intensity differences with X-ray or CT imaging techniques, as the brain mainly consists of soft tissues. Also, in contrast to X-ray imaging, MR images are recorded without damaging ionizing radiation. In addition, MR imaging can be used to study many different properties of the brain. Some of the MR imaging methods have become available only recently. For example, the brain activation detection and the study of connections between brain regions, to which Chapter 6 and Chapter 7 contribute.

In this thesis, newly developed methods, that contribute to the field of Neuro Imaging, are presented. The aim of these methods is to provide neurologists with improved tools that allow the study of the properties of, and processes in the (human) brain. As the author is no neurologist, the actual neurological interpretation will be limited in this thesis. However, in order to appreciate and understand the newly developed methods, some general facts

about the brain should be known:

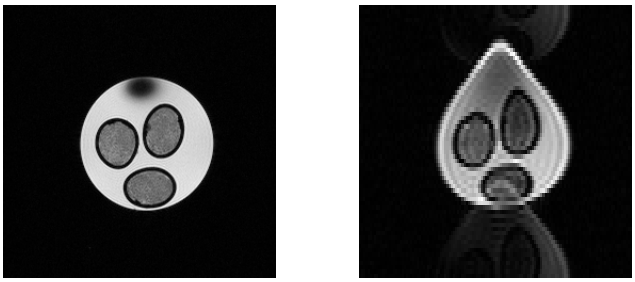
- The brain is the neural processing center that controls all actions and thoughts,
- The brain consists mainly of 2 tissue types, grey and white matter, and is surrounded by Cerebrospinal fluid (CSF). These different tissue types and the CSF are visible in Fig. 1.1.
- The grey matter is mainly located in the folded cortex at the outside of the brain, as well as in central structures such as the striatum, thalamus, amygdala, and hippocampus. The grey matter is the tissue where the neural processing, planning, and memory functions are located.
- The white matter is located more in the center of the brain and contains the connections ('phone cables') between brain regions. It consists mainly of (bundles of) myelinated axons which transport the action potential between different brain cells.

The brain has many more or less distinct cognitive functions, which are performed inside networks of the brain. In these networks, several grey matter structures are connected by white matter tracts. These white matter tracts are traditionally studied by diffusion weighted imaging (Chapter 7). Since the brain networks are distinct for different cognitive functions, the task depended grey matter activation has a different pattern for the different functions. These activation patterns are studied by functional MRI (fMRI) (Chapter 6). For some specific activations only a few grey matter regions are strongly involved, which allows easy separation of the activation patterns. For example, the activation pattern due to arm movement is very distinct from activation pattern due to visual stimuli. For other functions the separation is more difficult, thus requiring optimal grey matter activation detection. Although it is not part of the brain tissue, Cerebrospinal fluid (CSF) is also important for brain imaging. This fluid, which mostly consists of water, acts as a "cushion" or buffer for the brain. For most brain studies the CSF is not relevant, but it shows up in any brain image. Since it consists mainly of -homogeneous- water, the intensity of CSF might differ strongly from the intensity of brain tissue. These intensity differences might influence the analysis of nearby brain tissue due to effects such as Gibbs ringing.

1.3 Problems in MR imaging

1.3.1 Field inhomogeneities

A problem in MR imaging are inhomogeneities of the main magnetic field, the B_0 field. To obtain high quality MR images, the main magnetic field should



(a) SE image (b) EPI image

Figure 1.3: *A water filled cylinder, recorded by two acquisition methods, Spin Echo (SE) and Echo Planar Imaging (EPI). The SE method is much less sensitive to field inhomogeneities than the EPI method. In this image, the deformations are caused by an air bubble located just next to this slice.*

be constant to the high degree of approximately 1 part per million (ppm) over the entire imaged volume. However, the magnet is almost never constructed accurately enough to obtain a magnetic field that is constant to this precision. Furthermore, the magnetic properties of the object (head) also influences the magnetic field. For example, the relative permeability of water is 0.999 992 and of air it is 1.000 000 4, so there is a magnetic field inhomogeneity of 8 ppm when water and air are located next to each other. As different tissue types have different values for the permeability, and the location of the different tissues is not known a-priory, the resulting inhomogeneities in the magnetic field are not known a-priory. In practice, these magnetic field inhomogeneities are different for each different object and might depend on the orientation of the object with respect to the main magnetic field. As these magnetic field inhomogeneities are larger than allowed for some imaging modalities, especially Echo Planar Imaging (EPI), they will distort the recorded MR images. However, even with these distortions, EPI is a very popular acquisition method, since it is able to record images quickly. See Fig. 1.3, for an example of the distortions in an EPI image, caused by an air bubble in a water filled cylinder. The air bubble is located close to, but not in, the displayed slice. However, when the magnetic field inhomogeneities are known, it is possible to (partially) correct the distorted MR images. In Chapter 3, a method is developed to estimate the field inhomogeneities with a modified EPI sequence, in order to allow the correction of the distortions.

1.3.2 Long acquisition times or anisotropic MR images

With MR imaging, it is possible to record high resolution 3D images [8], with 2D or 3D sequences [9]. The difference between these is that 2D sequences

record the 2D k-space of a series of (adjacent) slices and 3D sequences record the 3D k-space of a volume. A problem with high resolution imaging is that it usually requires a long scan time. High resolution images require multiple radio excitations. However, there needs to be a certain time between repeated excitations of a volume (TR) and this minimum time depends on the type of acquisition. The TR is especially long when each new excitation of a specific volume requires T_1 relaxation in order to obtain the desired contrast, as the T_1 relaxation time is of the order seconds in most tissues. With 2D acquisitions, every excitation influences only a single slice, so within TR, a part of the k-space of all slices can be recorded. However, when the slices are very thin, the total signal power emitted by the slice is low and the excitation and recording of all of these individual slices will require longer than the minimal TR required for T_1 relaxation. This causes an extension of the total acquisition time, compared to an acquisition with thicker slices. Often, a compromise between acquisition speed and resolution is found in the recording of slices that are substantially thicker than the in plane resolution, i.e. these images have an anisotropic resolution. In this way, a good in plane resolution is combined with a modest/short acquisition time. The obvious disadvantage is the reduced resolution in the direction in which the slices are stacked. In Chapter 4, a method is developed that combines several of these anisotropic images into a single isotropic high resolution image.

1.3.3 Noise level estimation

In order to be able to discriminate between signal and noise when statistically analyzing MR images, it is necessary to know the noise level that is present in the images. Without the knowledge of the noise level, it is difficult to determine whether differences in the MR images are caused by differences in the object or by the noise that is present in the MR images. This is especially difficult when only one image is recorded, or when each image in a series of images has a different contrast. However, under some quite common conditions, the noise level in a MR image is constant over the entire image. Therefore, the level of the noise that is present in the entire image can be estimated from the background area, i.e. the area in which the object is not present. In Chapter 5, a method is presented with which the noise level of MR images is estimated by fitting the theoretical intensity distribution of the background intensities to the image histogram. In this method, the bins of the histogram that do contain the background mode are automatically selected.

1.3.4 Functional MRI

The aim of Functional MRI (fMRI) it to detect brain activity. In order to detect brain activity, images acquired with a method special for fMRI are recorded. Unfortunately, it is not possible to record MR images in which the brain activity

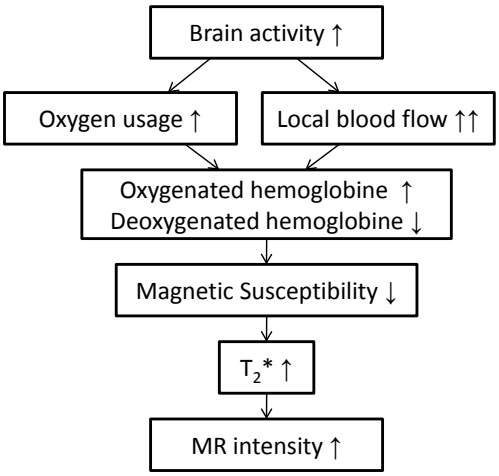


Figure 1.4: *This figure gives a schematic overview of the origin of the BOLD contrast, which is used for fMRI studies.*

is directly visible. Instead, the brain activity is indirectly measured through the influence it has on the oxygenation level of the nearby blood with Blood-Oxygen-Level Dependent (BOLD) sensitive MR images. Since the activation of specific brain areas is a dynamic process and changes relatively quickly, these BOLD sensitive images are usually acquired with fast imaging techniques such as EPI, which allow the acquisition of a time series of 3D images.

The BOLD contrast stems from the differences in magnetic properties of oxygenated and de-oxygenated hemoglobin, see Fig. 1.4. As the brain uses extra oxygen during activity, the concentration of oxygenated hemoglobin changes due to brain activity. However, the vascular system compensates, and even over-compensates, for the increased oxygen consumption. So, after a short delay due to the not-instantaneous vascular response, the concentration of oxygenated hemoglobin increases and the concentration of de-oxygenated hemoglobin decreases at locations with increased brain activity. This in turn reduces the magnetic susceptibility differences, which increases the T_2^* time constant, which increases the intensity of the BOLD sensitive MR images. Unfortunately, these brain activity induced MR intensity changes are low compared to the noise that is present in the MR images. Therefore, advanced statistical tests are needed to detect brain activation. These tests need to account for the delayed transition between the activity states, which is caused by delayed reaction of the vascular system. Also, these tests need to correctly model the noise that is present in the BOLD sensitive fMRI images. In Chapter 6, advanced tests to detect brain activity are developed, and the performance of these tests

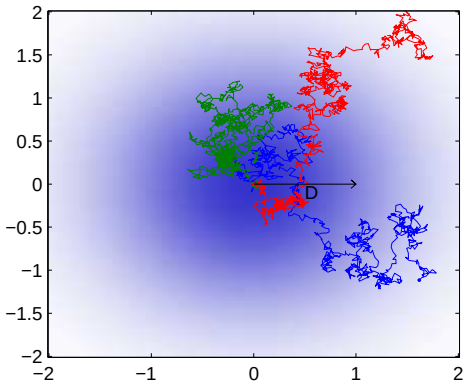


Figure 1.5: *Simulation of isotropic Brownian motion of three molecules, overlaid on the Probability Density Function (PDF) of the final displacement (shades of blue). The diffusion coefficient D, which is the expected diffusion distance, is indicated by the black arrow.*

is studied.

1.3.5 Diffusion weighted MRI

This section describes some problems that were encountered in diffusion weighted MRI. However, before introducing diffusion weighted MRI, first diffusion itself is briefly described. Diffusion is the random displacement of molecules due to Brownian motion, e.g. see Fig. 1.5. This Brownian motion is due to the thermal energy of the molecules. In isotropic media, such as pure water, the spatial displacement probability is given by a normal, also called Gaussian, distribution. However, the brain is not isotropic, especially in the white matter structures which consist of long white matter fibers. Therefore, the diffusion displacement is non isotropic and generally deviates from the normal distribution.

The diffusion process of water molecules can be measured with Diffusion Weighted MR images (DWI's), as DWI's are images that are made sensitive to diffusion of hydrogen nuclei. Even though these images depend on the diffusion of hydrogen nuclei, they mainly measure the diffusion of water, as water is the main liquid in the brain and each water molecule contains two hydrogen atoms. Diffusion weighted MRI is used to study the white matter of the brain, since the white matter consists of organized fibrous structures that have a micro structure in which the diffusion process is restricted in some directions. Along the direction of the white matter fibers, the water molecules can diffuse in a relatively unrestricted way. However, across the white matter fibers, the diffusion is restricted due to the presence of cell boundaries. Due to these

boundaries, which have low permeability, and different compartment sizes, the diffusion displacement deviates from the normal distribution.

The sensitivity of DWI's to the diffusion of hydrogen nuclei is generated by the application of a pair of diffusion weighting gradients. The first diffusion weighting gradient causes a de-phasing of the radio signal emitting nuclei and the second undoes this de-phasing, but only for nuclei that did not move in the time between the diffusion weighting gradients. For nuclei that did move between the application of these diffusion weighting gradients, the re-phasing is incomplete, resulting in destructive interference and thus a signal loss in those regions where diffusion in the direction of the applied diffusion weighting gradient is present. The amount of signal loss depends on both the diffusion in the direction of the diffusion weighting gradient and the strength of the diffusion weighting gradient.

In Chapter 7, a method is described with which the directions and the strength of the diffusion weighting gradients can be optimized for diffusion kurtosis imaging (DKI). DKI provides an extension to the more common diffusion tensor imaging (DTI) model, which assumes a Gaussian diffusion process. With DKI, this DTI model is extended to include the kurtosis, which is the 4^{th} order moment of the diffusion process.

1.4 Main contributions

The main contributions presented in this thesis are:

- A method to estimate the magnitude of field inhomogeneities (Chapter 3). The inhomogeneities of the main magnetic field are a main cause of distortions in Echo Planar Images (EPI). With the new non linear least squares estimator it is possible to estimate the field inhomogeneities from reference data with a substantial increased precision, compared to a previously developed method.

- A method to reconstruct high resolution MR images from a set of multi-slice MR images with anisotropic resolution (Chapter 4). This new method allows the reconstruction of a high resolution isotropic image from a series of multi slice MR images with thick slices, as long as the direction in which the slices are stacked is sufficiently different for each image.

- A method to estimate the noise level in magnitude MR images. This method uses a histogram of the image and computes the noise variance from the background mode (Chapter 5). When it can be assumed that the noise level is constant throughout the image, it is possible to estimate this noise level from the background area of the image. Since the background area is that area where the object does not contribute to the signal,

only the Rayleigh distributed noise is present in the background area of magnitude MR images. By fitting the Rayleigh distribution to the image histogram with a Maximum Likelihood estimator, the noise level is accurately determined.

- Advanced statistical tests for functional MRI activation detection were developed. The performance of these tests, in terms of sensitivity to activation and false alarm rate, is investigated (Chapter 6). The noise of the MR images used in functional MR, which studies brain activations, is not white but colored. This does influence the likelihood based statistical tests with which the brain activation can be detected. The actual effects of the coloring of the noise in finite length datasets are investigated and methods for correcting the false alarm rate, which is shown to deviate from the theoretical value, are presented.

- A method to optimize the diffusion weighting gradient settings for Diffusion Kurtosis Imaging (DKI) (Chapter 7). Since the DKI model is an extension of the DTI model to a higher order of the diffusion process, the parameters of the DKI are more sensitive to noise, compared to the parameters of the DTI model. The diffusion weighting settings with which a set of Diffusion Weighted Images (DWI) is acquired, influences the precision of the parameter estimates. Therefore, the variance of the parameters estimated from a set of DWI can be reduced by optimizing the diffusion weighting gradient settings of these DWI. The new optimization method finds the optimal diffusion weighting settings by minimizing the Cramér Rao lower bound of a suitable DKI parameter.

1.5 Manuscript Organization

This thesis is structured as follows:

Chapter 1 is this chapter, in which the work described in this thesis is briefly introduced.

Chapter 2 presents the methods that are used throughout the thesis.

Chapter 3 presents a method to estimate the strength of inhomogeneities in the main magnetic field of the MR machine. Since these inhomogeneities distort the MR images, they should be known with the highest possible accuracy and precision in order to be able to correct the distorted MR images.

Chapter 4 presents a method to create a high resolution MR image from several multi slice MR images with anisotropic resolution. The individual

MR images required by this method were acquired by a fast imaging sequence with different orientations of the slices.

Chapter 5 presents a method by which the noise level in MR images can be estimated. Knowledge of the noise level is essential for statistical analysis of MR images.

Chapter 6 presents several likelihood based functional MRI (fMRI) brain activation detection methods and studies their performance. Validating the optimal activation detection methods is essential for reliable interpretation of the detected brain activations.

Chapter 7 presents a method by which the diffusion weighting acquisition settings for Diffusion Kurtosis experiments can be optimized. With the optimized settings, the precision of Diffusion Kurtosis imaging (DKI) parameters is substantially improved.

Chapter 8 summarizes the conclusions of the work presented in this thesis.

Appendix A gives a description of the software that was developed in order to support the research presented in this thesis.

Chapter 2

Methods

This chapter introduces the methods that are used in the different chapters of this thesis. The first sections introduce important general concepts such as models and statistics, followed by a description of the statistical distributions with which the MR images can be described and the Maximum Likelihood estimator for MR images. Furthermore, advanced properties derived from the statistical distributions, such as the Fisher information and Cramér Rao lower bound, are introduced in the final section of this chapter.

2.1 Models

In this thesis, a model of the MR acquisitions is at the basis of each method. A model is a description of a system, or more specific, of measurements of a system. A model is a set of mathematical relations that describe, in our case, the MR images in terms of (interesting) parameters. With the aid of such a model, these parameters can be estimated from recorded MR images. A model tries to describe reality. However, all models are only an approximation of reality and can hardly ever describe all aspects perfectly. However, when properly validated, models can be used to obtain information about important aspects of the object under study.

2.2 Statistics

The methods that are developed in this thesis are based on the statistical properties of MR images.

Statistics describes the properties of random variables, i.e. it describes processes of which the outcome is random. A-priori, that is before the experiment has been performed, only the probability of each possible outcome can be specified. For example, the outcome of a single throw with a fair cubic dice can only be stated probabilistically: There is a chance of 1/6 to throw 6. A posteriori, i.e. after one or more experiments, the observed values might be used to validate the model (e.g. was it a fair dice?).

When the range of outcomes of a random variable is a set of real numbers, in contrast to the fixed set of integers of the dice, the random variable can be described by a probability density function (PDF) over the range of outcomes. This PDF might depend on some parameters. In this thesis, most PDF's describe the intensity of the MR images. Therefore, the PDF will depend on tissue properties such as the spin-lattice relaxation time (T_1), spin-spin relaxation time (T_2), BOLD, diffusion, and the acquisition parameters. The mathematical representation of the PDF, which depends on parameters, can be regarded as a statistical model.

2.3 Estimator

With measurements, i.e. the realizations of the random variables, it is possible to obtain information about the parameters of a statistical model. An estimator is a method to extract the information about the model parameters from the measurements. There exist several popular estimators, such as the least squares (LS) estimator and the maximum likelihood (ML) estimator. The LS estimator minimizes the sum of squared differences between the measured values and the values predicted by the model. The ML estimator maximizes the likelihood function, which is the probability density function where the observed values are substituted for the random variables. The ML estimator finds that value of the model parameters where the likelihood function is maximal. That is, the ML estimator finds those parameters for which the probability density of the observed measurements is maximal. When the PDF is a (multi variate) normal distribution, the ML estimator is equivalent to a (weighted) LS estimator. However, for PDF's with different distributions, the ML estimator and the LS estimator generally give different estimates. Whenever we measure, measurements unavoidably contain some noise. Thus, by measuring, it will be impossible to always obtain the true value of the parameters of the models. However, with good measurements and a good estimator, it is possible to obtain a good estimate of, i.e. approximation to, the true value of the parameters.

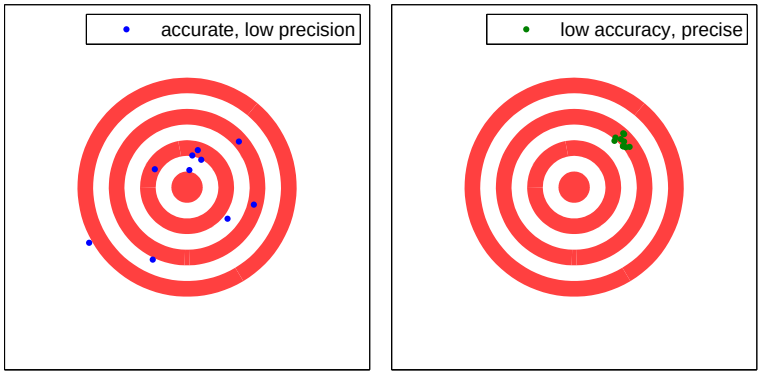


Figure 2.1: *This figure shows the difference between accuracy and precision when aiming for the center.*

2.4 Accuracy and precision

Important properties of estimators, which might also be used to compare different estimators, are the accuracy and precision, see Fig. 2.1. The accuracy measures how far an estimated value is 'on average' from the true value of the parameter in which we are interested. The distance of the 'average' to the true value is called the bias:

$$\text{bias}(X) = \mathbb{E}[X] - X_0, \tag{2.1}$$

where $\mathbb{E}$ is the expectation operator, X is the random variable that describes the outcome of the potentially biassed estimator, and X_0 is the true value of the process that we want to measure. Note that in general X_0 is unknown, except for simulation experiments. A good estimator has high accuracy and thus has a low, preferably zero, bias. The precision on the other hand is a measure of the 'average' spread of the measurements; roughly how much the value changes when the experiment is repeated. The variance, often used to quantify the precision, is the average of the square of this spread of the measured values:

$$\text{var}(X) = \mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big], \tag{2.2}$$

Obviously, a good estimator should also have high precision and thus a low variance. Note that, when noise is present in the measurements, it is impossible to obtain a variance of zero with an estimator that always has a low (zero) bias. These two properties, bias and variance, can be combined in the root mean square error (RMSE) measure:

$$\text{RMSE}(X) = \sqrt{\text{bias}(X)^2 + \text{var}(X)} = \sqrt{\mathbb{E}\Big[(X - X_0)^2\Big]}, \tag{2.3}$$

2.5 Distributions of MR images

In order to treat MR images in a statistical way, the distribution of the MR images needs to be known. In this section several distributions will be given, which are valid under different circumstances.

The noise that is present in the recorded MR images is caused by several different processes. A very prominent source of the noise are the receiver coils. This noise enters the complex k-space signal and is (almost) normal distributed, independent, and (almost) uncorrelated between the different k-space samples. As the Fourier Transform is a linear operation, the complex image $\tilde{y}$ is complex normally distributed as well:

$$\Re\left(\tilde{y}\right) \sim \mathcal{N}\left(\Re(\tilde{A}), \sigma\right) \tag{2.4}$$

$$\Im\left(\tilde{y}\right) \sim \mathcal{N}\left(\Im(\tilde{A}), \sigma\right), \tag{2.5}$$

where $\tilde{y}$ is the complex valued observation, $\Re$ and $\Im$ select the real($\Re$) and imaginary($\Im$) part of a complex value, $\tilde{A}$ is the 'true' complex valued MR signal without any noise contribution, and σ the standard deviation of the noise in the MR image. The probability density function (PDF) of $\tilde{y}$ is given by

$$p\left(\tilde{y}|\tilde{A}, \sigma\right) = \frac{1}{2\pi\sigma^2} e^{\frac{-1}{2\sigma^2}|\tilde{y}-\tilde{A}|^2}, \tag{2.6}$$

where $|.|$ computes the absolute value, or magnitude, of a complex value. The phase of the complex valued image $\tilde{y}$ is often considered to be not interesting for further analysis. Therefore, it is usually discarded and magnitude only images are stored. Unfortunately, the computation of the magnitude is a non-linear operation, so the noise distribution will be changed. It is well known that the magnitude operation changes the distribution into a Rician distribution [10,11]

$$y \sim \text{rice}(A, \sigma), \tag{2.7}$$

with $y = |\tilde{y}|$ and $A = |\tilde{A}|$. The PDF of the Rician distribution is given by

$$p\left(y|A, \sigma\right) = \frac{y}{\sigma^2} e^{\left(-\frac{y^2+A^2}{2\sigma^2}\right)} I_0\left(\frac{yA}{\sigma^2}\right) \epsilon(y), \tag{2.8}$$

where I_0 is the order zero modified Bessel function of the first kind and ϵ is the heavyside function, i.e. $\epsilon(y)$ is 1 for positive arguments and 0 for negative arguments.

The Signal to Noise Ratio (SNR), defined as A/σ, is zero in non-signal background areas. When the SNR is zero, it is known that the Rician PDF reduces to the Rayleigh PDF, which is given by:

$$p\left(y|\sigma\right) = \frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}} \epsilon(y) \quad . \tag{2.9}$$

This reduction of the Rician PDF to the Rayleigh PDF can be easily be proven with the asymptotic approximation of the ν^{th} order modified Bessel function:

$$I_\nu(z) \rightarrow \left(\frac{z}{2}\right)^\nu \Gamma(\nu+1) \quad \text{for} \quad z \rightarrow 0 \quad , \tag{2.10}$$

where Γ denotes the Gamma function. The moments of the Rayleigh PDF are given by:

$$\mathbb{E}\left[m^\nu\right] = (2\sigma^2)^{\nu/2}\Gamma\left(1+\frac{\nu}{2}\right) \quad , \tag{2.11}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator. The first and second moment of the Rayleigh distribution are often exploited to estimate the variance of background MR data [10, 12, 13].

When magnitude MR images are processed, for example by smoothing or even when interpolating in order to apply a spatial transformation, the distribution of the intensity changes. As far as we know, the distribution of linear combinations of rice distributed variables can not be given by an analytical expression, although with the central limit theorem it can be proven that this distribution approaches a normal distribution when the number of rice distributed values in the linear combination is large.

When multiple images are recorded in a series, either for fMRI, Chapter 6, or a set of DWI, Chapter 7, the correlation of the noise in this series of images is also important. In general, the time between the acquisition of these subsequent images is large compared to the typical correlation lengths present in the MR system. Therefore the MR system will not introduce any correlations into the series. However, when the activities of the brain are studied, unmodeled, arbitrary, spontaneous activation might be regarded as noise in the fMRI analysis. This noise source might have a non zero correlation between subsequent MR images. In Chapter 6 these correlations are explicitly modeled with an Auto Regressive (AR) model.

2.6 Maximum Likelihood estimator in MR

The aim of the Maximum Likelihood (ML) estimator is to estimate parameters of a model of MR images. For this, assume that there is a model $A_i(\boldsymbol{\theta})$ of the magnitude of a set of N MR magnitude images. Furthermore, assume that the magnitude MR images are not modified by smoothing, registration or interpolation and no correlation is present between the individual images in the image series, since otherwise the magnitude MR images are not rice distributed any more. This implies that any smoothing/interpolation and realignment should be done before computing the magnitude (i.e. it should be done in the complex domain). The model of the magnitude of the MR images can be any model that (tries to) describe the magnitudes of a specific voxel in the set of magnitude MR images, such as a fMRI model, see Chapter 6, DTI model, or DKI model, see Chapter 7. When the voxels of the images are

studied independently, spatial correlations that might be present in the MR images will not influence the validity of the following results.

The ML estimator optimizes the likelihood function, which is closely related to the PDF of the MR images. With the assumptions given above, the joint PDF of the measurements in a single voxel in the N magnitude MR images is given by

$$p(\boldsymbol{y}|\boldsymbol{A},\sigma) = \prod_{i=1}^{N} p(y_i|A_i,\sigma) \tag{2.12}$$

where $\boldsymbol{y}$ $(N \times 1)$ is a vector with the random variables that describe a specific voxel in all N images, $\boldsymbol{A}$ $(N \times 1)$ is a vector with the magnitudes predicted by the model for that same voxel, and p is the rice distribution, given in Eq. (2.8).

The ML estimator of $\boldsymbol{\theta}$ from N magnitude MR images can now in general be given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ln p(\boldsymbol{y}) = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \ln p(y_i), \tag{2.13}$$

in which $p(\boldsymbol{y})$ is the PDF given by Eq. (2.12), evaluated at the observed values of $\boldsymbol{y}$. In general the optimum cannot be found by an explicit analytical formula when rice distributions are involved, so nonlinear optimization techniques have to be used. Due to the similarity of the rice distribution to the normal distribution for large A/σ, a good initialization of the nonlinear minimization can usually be found by least squares (LS) estimation,

$$\hat{\boldsymbol{\theta}}_{LS} = \arg\min_{\boldsymbol{\theta}} |\boldsymbol{y} - \boldsymbol{A}|^2 . \tag{2.14}$$

The applicability of the LS initialization also depends on the actual model. Therefore, the specification of the initialization is given only after the model of the MR images is defined in the different chapters of this thesis.

2.6.1 Remarks for implementation

Non linear optimization techniques are often speeded up when the derivative of the function to be optimized is available. Therefore, the derivative of $\ln p(y_i)$ with respect to $\boldsymbol{\theta}$ is computed,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln p(y_i|A_i,\sigma) = \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\ln p(y_i|A_i,\sigma)}{\partial A_i} \tag{2.15}$$

$$= \left(\frac{-A_i}{\sigma^2} + \frac{y_i I_1\left(\frac{A_i y_i}{\sigma^2}\right)}{\sigma^2 I_0\left(\frac{A_i y_i}{\sigma^2}\right)} \right) \frac{\partial A_i}{\partial \boldsymbol{\theta}}, \tag{2.16}$$

in which (of course) the derivative of the model A with respect to the parameters $\boldsymbol{\theta}$ is still unknown.

2.7 Cramér Rao Lower Bound

The Cramér Rao Lower bound (CRLB) is a lower bound on the variance of parameters that are estimated with an unbiassed estimator. It can be proven that the maximum likelihood estimator is a consistent estimator, and thus reaches this lower bound asymptotically [14, 15]. Therefore, the CRLB is very useful for the study of the variance of the ML estimator. In this section, the general and model independent part of the CRLB of the model parameters $\boldsymbol{\theta}$ of the rice distributed MR images is given. The CRLB states that:

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}) \geq I(\boldsymbol{\theta})^{-1}, \tag{2.17}$$

where $I(\boldsymbol{\theta})$ is the Fisher information matrix. This information matrix is given by

$$I(\boldsymbol{\theta}) = \mathbb{E}\left[\left(\frac{\partial \ln p(\boldsymbol{y})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln p(\boldsymbol{y})}{\partial \boldsymbol{\theta}}\right)^{T}\right] \tag{2.18}$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^{N} \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(y_i)}{\partial A_i}\right)\left(\sum_{j=1}^{N} \frac{\partial A_j}{\partial \boldsymbol{\theta}^T} \frac{\partial \ln p(y_j)}{\partial A_j}\right)\right] \tag{2.19}$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial A_j}{\partial \boldsymbol{\theta}^T} \mathbb{E}\left[\frac{\partial \ln p(y_i)}{\partial A_i} \frac{\partial \ln p(y_j)}{\partial A_j}\right]. \tag{2.20}$$

Obviously, the derivatives of A_i with respect to $\boldsymbol{\theta}$ will depend on the actual model A and will not be discussed further in this section. However, it is possible to expand the expectation term of Eq. (2.20):

$$\mathbb{E}\left[\frac{\partial \ln p(y_i)}{\partial A_i} \frac{\partial \ln p(y_j)}{\partial A_j}\right] = \iint_{y_i, y_j = 0}^{\infty} p(y_i) p(y_j) \frac{\partial \ln p(y_i)}{\partial A_i} \frac{\partial \ln p(y_j)}{\partial A_j} \, dy_i \, dy_j. \tag{2.21}$$

For $i \neq j$ this is equal to:

$$= \int_{y_i=0}^{\infty} p(y_i) \frac{\partial \ln p(y_i)}{\partial A_i} \, dy_i \int_{y_j=0}^{\infty} p(y_j) \frac{\partial \ln p(y_j)}{\partial A_j} \, dy_j. \tag{2.22}$$

The two terms in this product can be expanded to:

$$\int_{y_i=0}^{\infty} p(y_i) \frac{\partial \ln p(y_i)}{\partial A_i} \, dy_i \tag{2.23}$$

$$= \int_{y_i=0}^{\infty} \frac{y_i}{\sigma^2} e^{\left(-\frac{y_i^2+A_i^2}{2\sigma^2}\right)} I_0\left(\frac{y_i A_i}{\sigma^2}\right) \left(\frac{-A_i}{\sigma^2} + \frac{y_i I_1\left(\frac{A_i y_i}{\sigma^2}\right)}{\sigma^2 I_0\left(\frac{A_i y_i}{\sigma^2}\right)}\right) dy_i \tag{2.24}$$

$$= \frac{-A_i e^{\frac{-A_i^2}{2\sigma^2}}}{\sigma^4} \int_{y_i=0}^{\infty} y_i e^{\frac{-y_i^2}{2\sigma^2}} I_0\left(\frac{y_i A_i}{\sigma^2}\right) dy_i + \frac{e^{\frac{-A_i^2}{2\sigma^2}}}{\sigma^4} \int_{y_i=0}^{\infty} y_i^2 e^{\frac{-y_i^2}{2\sigma^2}} I_1\left(\frac{y_i A_i}{\sigma^2}\right) dy_i \tag{2.25}$$

$$= \frac{-A_i e^{\frac{-A_i^2}{2\sigma^2}}}{\sigma^4} \sigma^2 e^{\frac{A_i^2}{\sigma^2}} + \frac{e^{\frac{-A_i^2}{2\sigma^2}}}{\sigma^4} \sigma^2 A_i e^{\frac{A_i^2}{\sigma^2}} \tag{2.26}$$

$$= 0 \quad , \tag{2.27}$$

where the following (simplified) identities from [**?**] were used:

$$\frac{1}{t}\frac{\partial}{\partial t} t^{-\nu} I_\nu(t) = t^{-\nu-1} I_{\nu+1}(t) \tag{2.28}$$

$$\int_{y_i=0}^{\infty} t^{\nu+1} e^{-a^2 t^2} J_\nu(bt) \, dt = \frac{b^\nu}{(2a^2)^{\nu+1}} e^{\frac{b^2}{4a^2}} \tag{2.29}$$

$$I_\nu(t) = \mathrm{i}^{-\nu} J_\nu(\mathrm{i}t), \tag{2.30}$$

where J_ν is the Bessel function of the first kind. For $i = j$ in Eq. (2.21), the expectation term is the Fisher information of a rice distributed variable, $I_{\text{rice}}(A, \sigma)$, which can be simplified to

$$I_{\text{rice}}(A_i, \sigma) = \mathbb{E}\left[\left(\frac{\partial \ln p(y_i)}{\partial A_i}\right)^2\right] = \int_{y_i=0}^{\infty} p(y_i) \left(\frac{\partial \ln p(y_i)}{\partial A_i}\right)^2 dy_i \tag{2.31}$$

$$= \int_{y_i=0}^{\infty} \frac{y_i}{\sigma^2} e^{\left(-\frac{y_i^2+A_i^2}{2\sigma^2}\right)} I_0\left(\frac{y_i A_i}{\sigma^2}\right) \left(\frac{-A_i}{\sigma^2} + \frac{y_i I_1\left(\frac{A_i y_i}{\sigma^2}\right)}{\sigma^2 I_0\left(\frac{A_i y_i}{\sigma^2}\right)}\right)^2 dy_i. \tag{2.32}$$

This integral cannot be analytically solved, but it can be numerically computed and/or tabulated. When this rice information is substituted in Eq. (2.20), the Fisher information matrix of the parameters $\boldsymbol{\theta}$ is simplified to

$$I(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^T} I_{\text{rice}}(A_i, \sigma). \tag{2.33}$$

Note that in order to compute the Fisher information with Eq. (2.33), the magnitudes predicted by the model, A_i, and the derivatives of this magnitude with respect to the model parameters, $\frac{\partial A_i}{\partial \boldsymbol{\theta}}$, need to be available.

2.7.1 Remarks for implementation

As noted, the integral Eq. (2.32) cannot be analytically solved, but it essentially only depends on A_i/σ, as is demonstrated by introducing the substitutions $a = A_i/\sigma$ and $\bar{y} = y_i/\sigma$,

$$I_{\text{rice}}(a\sigma, \sigma) = \frac{1}{\sigma^2} \int_{\bar{y}=0}^{\infty} \bar{y} e^{\frac{-1}{2}(\bar{y}^2+a^2)} I_0(\bar{y}a) \left(-a + \frac{\bar{y} I_1(\bar{y}a)}{I_0(\bar{y}a)} \right)^2 d\bar{y}. \tag{2.34}$$

As the integral in Eq. (2.34) only depends on a, it can be computed by numerical integration and tabulated for reuse. Our implementation treats small and large a separately. For $a \leq \frac{326787}{131072} \approx 2.5$, the value of I_{rice} is computed by one of the 87 precomputed 6^{th} order interpolating polynomials in a. Each precomputed polynomial is valid on a small domain. For $a > \frac{326787}{131072}$, the value of I_{rice} is computed by one of the 27 precomputed 6^{th} order interpolating polynomials in $1/a^2$. The resulting value is accurate to full double precision over the entire domain of positive real numbers, as the residual error is approximately 2^{-52}. The cutoff value, the domains, and order of the interpolating polynomials are chosen to minimize table size and computation time.

Note that when a normal distribution is assumed for the measurements, the equivalent Fisher information $I_{\text{normal}} = \frac{1}{\sigma^2}$. Furthermore, it is easy to show that the asymptotic behavior of Eq. (2.34) is given by:

$$I_{\text{rice}} = \frac{1}{\sigma^2} \begin{cases} a^2 & \text{for} \quad a \ll 1 \\ 1 - \frac{1}{2a^2} & \text{for} \quad a \gg 1 \end{cases} \tag{2.35}$$

Chapter 3

Improved B_0 field map estimation for high field EPI

The major part of the work described in this chapter is published as:

Dirk H. J. Poot, Wouter Pintjens, Marleen Verhoye, Annemie Van Der Linden and Jan Sijbers

Improved B_0 field map estimation for high field EPI

in: Magnetic Resonance Imaging,
Volume 28, Issue 3, April 2010, Pages 441-450

Abstract

Echo Planar Imaging (EPI) is an ultrafast Magnetic Resonance Imaging (MRI) technique that allows one to acquire a 2D image in about 100ms. Unfortunately, the standard EPI images suffer from substantial geometric distortions, mainly originating from susceptibility differences in adjacent tissues. To reduce EPI distortions, correction methods based on a field map, which is a map of the off-resonance frequencies, have been developed. In this chapter, a *nonlinear least squares estimator* is used to optimize the estimation of the field map of the B_0 field. The model of the EPI and reference data includes parameters for the phase evolution, the complex magnitude, the relaxation of the MRI signal, and the EPI-specific phase difference between odd and even echoes. With these parameters additional corrections of MR images might be computed. The reference data required to estimate the field map can be

acquired with a modified EPI-sequence.
The proposed method is tested on simulated as well as experimental data and proves to be significantly more robust against noise, compared to the previously suggested method.

keywords Field mapping, Parameter Estimation, Susceptibility artifacts, Echo Planar Imaging corrections

3.1 Introduction

Echo Planar Imaging (EPI) [16] is an ultrafast imaging technique, well-suited for MRI applications that require high temporal resolution (e.g., functional Magnetic Resonance Imaging (fMRI)) or in which a large number of different images of the same object have to be acquired (e.g. Diffusion Tensor Imaging (DTI) [17,18]). The main drawback of EPI is its sensitivity to off resonance factors such as B_0 field inhomogeneity, chemical shifts, and eddy current effects from fast switching gradients. These effects introduce image artifacts, especially at high fields (7T and higher). Since high field scanners are common in small animal imaging and are also starting to enter clinical applications, it is of major importance that correction strategies for EPI distortions are developed.

In the past, various methods to correct EPI distortions methods were proposed. These methods can be subdivided in three categories:

acquisition The first category of EPI distortion correction methods are methods that are applied at the hardware level during the acquisition. These methods commonly use shim coils, which generate spherical harmonic magnetic fields, to compensate for global and local field inhomogeneities. Conventional global shimming techniques try to optimize the field homogeneity for the entire imaged volume [19]. However, since the order of the spherical harmonic magnetic fields generated by the shim coils is limited, they cannot correct for all susceptibility differences in the brain. Dynamic shimming [20] has been shown to improve magnetic field homogeneity to a larger extent than conventional global shimming, since it optimizes the homogeneity of the main magnetic field by updating the shim settings for each slice separately. However, a drawback of dynamic shimming is the high performance of the shim coils that is required.

registration A second method to correct EPI distortions employs image registration [21] and post processing. For this method, a distortion free reference image, such as a Spin Echo (SE) image, is required. The EPI images are then registered to this reference image. Unfortunately, since field inhomogeneities cause local distortions, it is generally not possible to perform adequate corrections with affine transformations of the image

only. Hence, more advanced, non-affine registration techniques are required. A drawback of these methods is that there might be insufficient contrast for the registration or the reference image might be of a different modality, which complicates the registration.

field mapping A last category of EPI distortion correction methods combines a special acquisition and post processing techniques. For these methods, extra reference data are acquired. From the reference data, a deformation field can be computed. This deformation field can then be used by post processing techniques to undo the deformations in the EPI images. Previously proposed methods that employ this approach measure the main field inhomogeneities [22–25], or the Point Spread Function (PSF) [25,26].

In this chapter, a technique from the last category, the field mapping technique, is optimized. For the standard field mapping [22], at least two images with different echo times have to be acquired. From the phase difference between these images, a field inhomogeneity map or off-resonance frequency map is calculated. To obtain reliable displacement maps, the standard field mapping requires the unwrapping of the phase discontinuities. A refinement of this method, which avoids the need for phase unwrapping, was proposed by Schmithorst et al. [23], who acquired multiple gradient echo (GRE) images to estimate the field map. However, due to the differences between even and odd EPI echoes, this method requires the reference data to be divided in two parts, containing either the echoes with odd or even echo number. Moreover, relaxation effects were not taken into account.

We propose a new field mapping technique based on an improved model of the reference data along with a nonlinear least squares estimator. The model parameters represent properties of the MRI recording, the complex amplitude, the off-resonance frequency, the T_2^* relaxation, and the variation between even and odd EPI echoes. The proposed model does not require the splitting of the data in parts containing only the even and odd echoes prior to the estimation of the parameters. The performance in terms of the root mean square error (RMSE) and bias of this method is investigated by simulation and real data experiments. These experiments test the robustness to noise as well as the amount of reference data needed.

3.2 Methods

3.2.1 Field mapping

In magnetic resonance imaging, the demodulated MR signal $S(t)$ of an excited volume Ω, generated by freely precessing nuclear spins in the presence of a linear magnetic field gradient $\boldsymbol{G}$, equals the Fourier transform of the effective density $\rho(\boldsymbol{r})$. This effective density is the proton density weighted by the relevant

decays (T1,T2,...) and contrasts (e.g. diffusion weighting), and in general, it is complex valued. When no distorting effects are present, the recorded signal is given by [27]

$$S(t) = \iint_{\boldsymbol{r}\in\Omega} \rho(\boldsymbol{r}) e^{-\mathrm{i}\boldsymbol{k}(t).\boldsymbol{r}}\, d\boldsymbol{r} \quad . \tag{3.1}$$

In Eq. (3.1), $\boldsymbol{k}$ is a vector in k-space of which the components are given by

$$k_j(t) = \gamma \int_0^t \boldsymbol{G}(t').\boldsymbol{e}_j\, dt' \quad , \tag{3.2}$$

with $\boldsymbol{G}(t')$ the applied gradient at time t', γ the gyromagnetic ratio, and $\boldsymbol{e}_j$ the cartesian unit vector in the direction j. From Eq. (3.1), it is clear that the image reconstruction involves an inverse Fourier transform, which can efficiently be computed when the signal is sampled on a regular grid in k-space, which it is for echo planar imaging (EPI), at least when no samples are recorded during the gradient switching.

In Eq. (3.1), it is assumed that only the gradients $\boldsymbol{G}$ affect the acquired MR signal. However, in practice, additional factors such as timing offsets t_{off}, susceptibility effects causing an extra off-resonance frequency term $\omega(\boldsymbol{r})$, and T_2^* decay, affect the measured signal $S(t)$. Including these effects in Eq. (3.1) yields a more realistic model of the acquired MR signal:

$$\tilde{S}(t) = \iint_{\boldsymbol{r}\in\Omega} \rho(\boldsymbol{r}) e^{-\mathrm{i}[\boldsymbol{k}(t+t_{off})\boldsymbol{r} - \omega(\boldsymbol{r})t] - \frac{1}{T_2^*}t}\, d\boldsymbol{r}. \tag{3.3}$$

In a conventional EPI acquisition scheme the k-space is sampled line by line after one excitation. Hence, the off-resonance frequency $\omega(\boldsymbol{r})$ generally leads to a shift of the reconstructed position of $\rho(\boldsymbol{r})$, in the phase encoding direction. Since the field inhomogeneities, and therefore ω, are, by definition, not constant in Ω, these field inhomogeneities will cause geometric distortions. By using reference data, $\omega(\boldsymbol{r})$ can be estimated and the geometric distortions can be corrected. In the next subsection, we will describe how the reference data is acquired.

3.2.2 Reference data

Reference data is acquired with an adjusted EPI sequence as shown in Fig. 3.1a. A standard EPI phase encoding scheme with N gradient echoes records a different k-space line with each echo. The echoes originate from the alternating amplitude of the read-out gradient, and the phase encoding gradient is used to select the line in k-space. However, in the sequence for the reference data acquisition, the EPI phase encoding scheme is replaced by the phase encoding of a conventional GRE sequence [23], so a k-space line is sampled N times after an excitation pulse.
When all k-space lines for each echo number are combined, N images can be

reconstructed by Fourier transforming each *read-phase* plane of the data cube. Each image j (with $j = 0, ..., N-1$) has a different echo time $t_j = t_0 + jT_r$, where t_0 is the time between the radio pulse and the center of the first echo and T_r is the time between two subsequent gradient echoes, see (Fig. 3.1b). The differences between these images recorded with different echo times are caused by relaxation, odd/even phase shift, and the off-resonance frequency $\omega(\boldsymbol{r})$. When T_r is small enough to ignore relaxation and field inhomogeneity effects during the readout of a single line, the model of the FFT reconstructed GRE images I is given by

$$I(\boldsymbol{r}, j) = \rho(\boldsymbol{r}) e^{(\mathrm{i}\omega(\boldsymbol{r}) - \frac{1}{T_2^*(\boldsymbol{r})})t_j + i\varphi(\boldsymbol{r})\mathrm{mod}(j,2)}, \tag{3.4}$$

where $\boldsymbol{r}$ is the position in the plane, $\varphi(\boldsymbol{r})$ is the phase difference between the even and the odd images, caused by t_{off} and $\mathrm{mod}(j, 2)$ computes j modulo 2, which is zero for even j and 1 for odd j. Although in practice decay of transverse magnetization may be more complex than reflected in the mono-exponential form of Eq. (3.4), this model is expected to be sufficiently accurate, since our aim is to estimate the phase trend $\omega(\boldsymbol{r})$ from the reference data. The magnitude and phase are orthogonal directions, coupled mainly by the magnitude dependence of the phase variance. I.e. the variance of the phase depends on the magnitude, but not the actual phase value. Therefore, small errors in the magnitude model will not strongly influence the phase (trend) estimates.

In the next section, an existing method to estimate $\omega(\boldsymbol{r})$ as well as a new method to estimate $\rho(\boldsymbol{r})$, $T_2^*(\boldsymbol{r})$, $\omega(\boldsymbol{r})$, and $\varphi(\boldsymbol{r})$ from the reference data will be described.

3.2.3 Autocorrelation method

The *phase correction method* (CORR) in [23], a modified version of the method in [28], uses the autocorrelation function R to estimate the field map. The autocorrelation of a series of N_1 complex values z_j ($j = 0, \ldots, N_1 - 1$), without subtracting the mean, is given by

$$R(m) = \begin{cases} \sum_{j=0}^{N_1-1-m} (z_{j+m})(z_j)^* & m \geq 0 \\ R^*(-m) & m < 0 \end{cases}. \tag{3.5}$$

Due to the even-odd echo asymmetry, the even and odd echo images are processed separately:

$$z_{even,j}(\boldsymbol{r}) = I(\boldsymbol{r}, 2j) \tag{3.6a}$$

$$z_{odd,j}(\boldsymbol{r}) = I(\boldsymbol{r}, 2j+1), \tag{3.6b}$$

where $j = 0, \ldots, N_1$, with $N_1 = \lfloor N/2 \rfloor$ for z_{even} and $N_1 = \lceil N/2 \rceil$ for z_{odd}. From these two time series, for each voxel, R_{even} and R_{odd} are computed,

where the position argument $\boldsymbol{r}$ is not shown to simplify notation. The phase trend is present in $\Phi\left(R(1)\right)$, where Φ returns the phase of a complex value. The estimator of the off-resonance frequency $\omega(\boldsymbol{r})$ is then given by

$$\hat{\omega}_{\mathrm{CORR}}(\boldsymbol{r}) = \frac{\Phi\left(R_{even}(1)\right) + \Phi\left(R_{odd}(1)\right)}{4T_r}. \tag{3.7}$$

Note that this procedure does not account for relaxation.

3.2.4 Nonlinear least squares estimator

Our proposed *nonlinear least squares* (NLLS) method to estimate $\omega(\boldsymbol{r})$ is based on the complex valued data model from Eq. (3.4). In order to use real-valued optimization routines, which are most common, the function is re-parameterized for each position $\boldsymbol{r}$ as

$$f(j, \boldsymbol{\lambda}(\boldsymbol{r})) = e^{\mathrm{i}[\lambda_1 j + \lambda_2 + \lambda_3 \mathrm{mod}(j,2)] + \lambda_4 j + \lambda_5}, \tag{3.8}$$

with

$$\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_5] \tag{3.9}$$

$$= \left[\omega(\boldsymbol{r})T_r, \Im\left\{\ln \rho(\boldsymbol{r})\right\}, \varphi, -\frac{T_r}{T_2^*}, \Re\left\{\ln \rho(\boldsymbol{r})\right\}\right], \tag{3.10}$$

where $\Re\{\}$ and $\Im\{\}$ return the real and imaginary part of a complex value, respectively. For each position $\boldsymbol{r}$, the function $f(j, \boldsymbol{\lambda})$ is fitted to the N data points $I(\boldsymbol{r}, j)$ in a least squares sense, with respect to $\boldsymbol{\lambda}$:

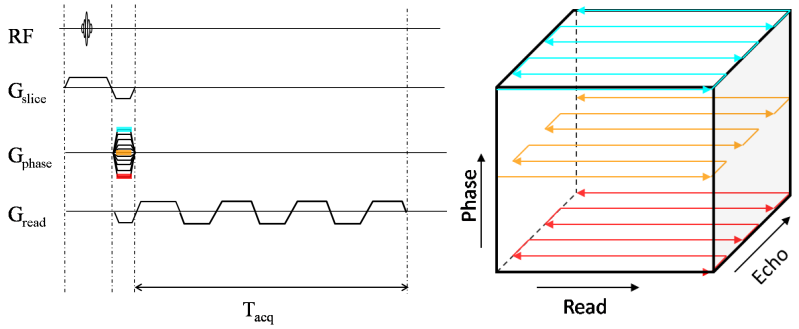
$$\hat{\boldsymbol{\lambda}}(\boldsymbol{r}) = \arg\min_{\boldsymbol{\lambda} \in \mathbb{R}^5} \sum_{j=0}^{N-1} \left| f(j, \boldsymbol{\lambda}) - I(\boldsymbol{r}, j) \right|^2, \tag{3.11}$$

where the $\hat{}$ indicates an estimated value. The NLLS estimate of the off-resonance frequency $\omega(\boldsymbol{r})$ is then given by

$$\hat{\omega}_{\mathrm{NLLS}}(\boldsymbol{r}) = \frac{\hat{\lambda}_1}{T_r} \quad . \tag{3.12}$$

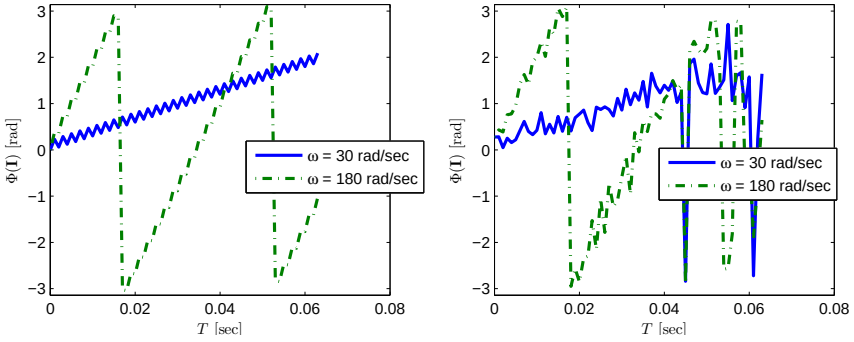
During the optimization, no constraints are applied. However, due to the periodicity of the exponential function in Eq. (3.8), the resulting parameter vector estimate $\hat{\boldsymbol{\lambda}}(\boldsymbol{r})$ given in Eq. (3.11), can always be mapped to satisfy $|\lambda_1| \leq \frac{\pi}{2}$, $|\lambda_2| \leq \pi$, and $|\lambda_3| \leq \pi$. Note that the correlation estimator in Eq. (3.7) produces phase trend estimates $\hat{\omega} T_r$ in the interval $[-\frac{\pi}{2}, +\frac{\pi}{2}]$ as well.

As long as individual parts of the object are displaced by less than half the field of view in an EPI image, no phase unwrapping of $\hat{\lambda}_1$ (and indirectly $\hat{\omega}_{\mathrm{NLLS}}$) is needed. This is usually ensured, as any MR imaging modality is influenced by field inhomogeneities that are so strong.



(a) Acquisition sequence (b) Data cube

Figure 3.1: *(a) The sequence used for measuring the field map is a conventional EPI readout train, but the phase encoding gradient is replaced by the phase encoding gradient of a GRE sequence. (b) The data from the field map sequence is shown on a data cube. The color of the EPI train corresponds to the color of the selected phase encoding step from (a).*



(a) Phase, no noise (b) Phase, $\sigma = 0.2$

Figure 3.2: *One realisation of the phase of the simulated signals with both phase trends. No noise was added in figure (a), noise with $\sigma = 0.2$ was added in figure (b).*

3.3 Experiments

Simulation as well as imaging experiments were run to compare the *phase correction method* CORR with the proposed *nonlinear least squares* (NLLS) method in terms of the precision and accuracy of the field map estimation. Reference data was simulated and the field map $\omega(\boldsymbol{r})$ was estimated with both methods. In addition, to test the performance of the estimators with real data, the different field map estimators were compared on experimental EPI images along with reference data.

3.3.1 Simulation experiments

In order to test the precision and accuracy of the phase trend estimators as a function of the number of echoes N, the noise standard deviation σ, and the phase trend magnitude ω, several simulation experiments were performed. To this end, two reference data sets were simulated with $T_2^* = 30\text{ms}$, $T_r = 1\text{ms}$, $\rho = 2$, and $\varphi = 0.2\text{rad}$, which were held constant throughout all the simulation experiments. The first reference data set was simulated with $\omega = 30\,\text{rad/sec}$, which did not cause a phase jump and the second reference data set was simulated with $\omega = 180\,\text{rad/sec}$, which caused two phase jumps. Fig. 3.2a shows the phase of both signals without noise added and with $N = 64$, and Fig. 3.2b shows one simulation of both signals after Gaussian noise with $\sigma = 0.2$ was added. The value for ρ and σ should be interpreted in terms of the SNR, which is given by

$$\text{SNR} = \frac{P_{signal}}{P_{noise}} = \frac{|\rho|^2 \left(1 - e^{-2N\,T_r/T_2^*}\right)}{2\sigma^2 N \left(1 - e^{-2T_r/T_2^*}\right)} \quad , \tag{3.13}$$

where P_{signal} and P_{noise} denote the power of the signal and noise, respectively.

During all simulation experiments, the root-mean-squared-error (RMSE) and the bias of the CORR and NLLS field map estimators were analyzed.

Three Monte Carlo simulations were produced according to the following protocols:

- The first simulation experiment tested the precision and accuracy of the phase trend estimators as a function of σ. For this, independent gaussian noise with standard deviation $0 \leq \sigma \leq 0.9$ was added to the real and imaginary parts. In this experiment, the number of echoes was held constant at $N = 64$ and the number of Monte Carlo realizations for each tested σ was $M = 100\,000$.

- The second simulation experiment investigated the effect of changing the number of echoes $3 \leq N \leq 100$. In this simulation experiment, the noise level was fixed to $\sigma = 0.2$ and the number of Monte Carlo realizations for each value of N was $M = 100\,000$. Since the signal decays, the Signal to

Noise Ratio (SNR) will depend on the number of echoes N. The CORR and NLLS methods were again used to estimate ω.

- In the third simulation experiment, both the number of echoes N and the standard deviation σ were varied, where $3 \leq N \leq 100$ and $0 \leq \sigma \leq 0.9$ and the number of Monte Carlo realizations for each combination of N and σ was $M = 10\,000$.

3.3.2 Imaging experiments

In order to investigate the performance of the methods on real data, three different datasets of a DTI hardware phantom were acquired with a 7T Pharmascan small animal system, manufactured by Bruker (Ettlingen, Germany). The DTI hardware phantom consists of parallel bundles of woven strands of Micro Dyneema fibers [29]. The first set was a DTI dataset, which was recorded with an EPI sequence ($TE = 35$ ms, $TR = 3000$ ms, imaging matrix $= 128 \times 64$). This dataset contained substantial geometric distortions due to susceptibility artifacts (see Fig. 3.7b).

To enable the correction of these geometric distortions, reference data, as described in subsection 3.2.2, were recorded with the same parameters. To be able to compare the quality of the estimated field maps, the acquisition of the reference data was repeated 10 times, with an artificially increased noise level to more clearly identify the effects of the noise on the estimated field maps. For these reference datasets, the number of echoes N in the multi echo gradient echo (GRE) sequence equals the number of phase encoding steps of the EPI sequence. Hence, for each reference dataset, $N = 64$ GRE images with different echo times were acquired.

To validate the correction results, a 256×128 Spin Echo (SE) image was recorded with $TE = 43$ms, $TR = 1500$ ms. Since an SE sequence is less sensitive to susceptibility artifacts than EPI, the image recorded with this SE sequence can serve as a suitable basis for comparison of the corrected EPI image. For all data sets, 20 slices of 1 mm thickness were acquired and the field of view (FOV) was 45 mm.

3.3.3 Implementation details

For all simulations and experiments presented in this chapter, MATLAB (The MathWorks, Inc. Natick, MA, USA) was used with custom routines. The optimization of the NLLS method used the standard nonlinear least squares routine (`lsqnonlin`). This routine is a local optimization routine, and thus there is no guarantee that the global minimum will be found. However, when the initial values are sufficiently close to the position of the global minimum, the routine will converge to that. In Appendix 3.A a Fourier based initialization of $\boldsymbol{\lambda}$ is described. This method was used in the remainder of this chapter and

it was observed that with this initialization the global minimum was almost always found, especially for low σ. Furthermore, a good initialization of $\boldsymbol{\lambda}$ will decrease the number of iterations needed to reach the optimum. At our machine (2.4Ghz Intel Core 2 Quad CPU), the initialization and estimation procedure took approximately 9.6 ms per voxel.

3.4 Results and discussion

3.4.1 Simulation experiments

This section discusses the results of the simulation experiments described in subsection 3.3.1.

Performance as a function of the noise level

Fig. 3.3 shows the RMSE as well as the bias of the field map estimators as a function of σ for the first data set ($\omega = 30$ rad/sec) with $N = 64$. The lines in the figures indicate the observed values and the shaded areas represent the 95% confidence intervals. Fig. 3.3a shows that the bias of the CORR and the NLLS estimator cannot be proven to be non-zero in this simulation.

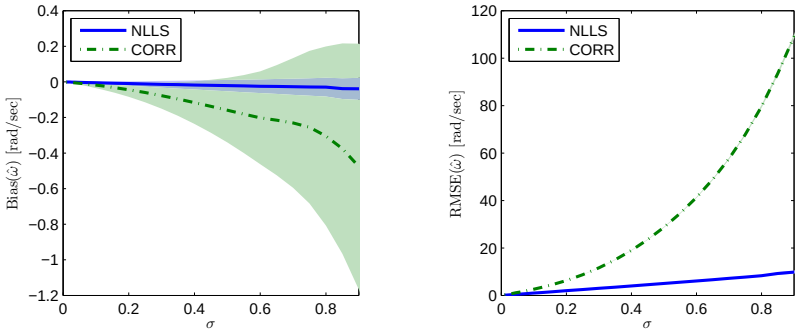
Fig. 3.3b shows that the RMSE of the NLLS estimator is substantially smaller than the RMSE of the correlation estimator for all noise levels. Note that the RMSE of both estimators is mainly caused by the variance of the estimators, not by the bias. Hence, the increase of the RMSE visible in Fig. 3.3b is mainly due to the increasing noise level.

The scaling of the RMSE as a function of σ obscures the relative performance of the different estimators. Therefore, to compensate for the expected relation between RMSE and noise level, Fig. 3.4a shows the RMSE scaled by $1/\sigma$. Note that, for the NLLS estimator, the scaled RMSE is constant, which indicates constant efficiency of the estimation of the field map by this estimator. On the other hand, the scaled RMSE of the CORR estimator increases with increasing σ, which indicates that the estimator becomes less efficient with increasing noise level.

Fig. 3.4b shows the results when phase jumps are present in the data. Comparison of Fig. 3.4a and Fig. 3.4b shows that the RMSE of CORR and NLLS estimators are not significantly influenced by the phase jump.

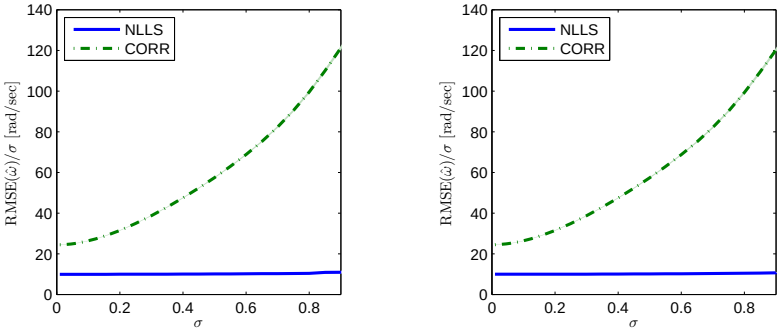
Performance as a function of the number of gradient echoes

Fig. 3.5 shows the performance of the field map estimators as a function of N. Fig. 3.5a shows the RMSE while Fig. 3.5b shows this RMSE scaled by $\sqrt{N}$ to remove the main trend of the RMSE. As can be seen in Fig. 3.5b, the RMSE of the estimators sharply decreases with N, when N is small. Then, it levels off and for large N the scaled RMSE starts to increase again. Note that the



(a) Bias (b) RMSE

Figure 3.3: *The bias (a) and RMSE (b) of the CORR and the NLLS field map estimators from simulated data with $\omega = 30$ rad/sec and $N = 64$. The shaded areas represent the 95% confidence regions of the performance measures. For these simulations, $M = 100\,000$ realizations were used.*



(a) (b)

Figure 3.4: *This figure shows the RMSE/σ of the different methods for the simulated signals with $\omega = 30$ rad/sec (a) or $\omega = 180$ rad/sec (b). Also for these results $N = 64$, $M = 100\,000$ and the shaded areas represents the 95% confidence regions of the scaled RMSE. These figures show the RMSE divided by σ to demonstrate the difference between the methods more clearly.*

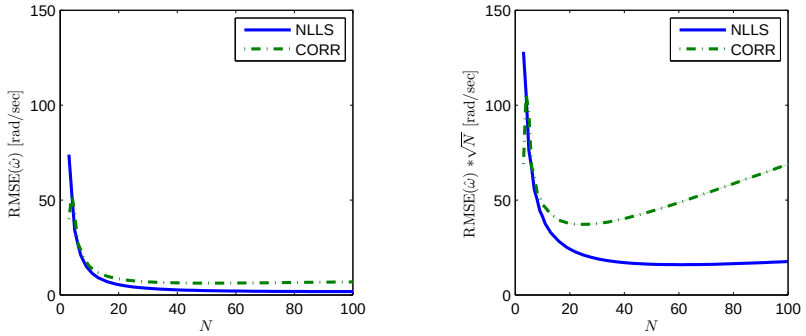
RMSE itself, however, does not increase, even for large N. This decrease and increase of scaled RMSE is expected. First, the scaled RMSE decreases with increasing number of echoes, since the linear trend of the phase of the simulated series of echoes is estimated. To accurately estimate a linear trend, the samples should be separated by as large a distance as possible. Therefore, increasing the maximum distance between the samples by adding an extra sample (i.e. record an extra echo) decreases the (scaled) RMSE. Secondly, when the number of echoes is increased beyond a certain limit, the scaled RMSE increases. This is also expected, since the magnitude of ρ is fixed and each subsequent echo has a lower magnitude due to the T_2^* relaxation. Beyond a certain number of echoes, the magnitude will be so low that the amount of information added by each subsequent echo is less than expected by the scaling, which assumes a constant amount of information per echo.

Performance as a function of noise level and number of echoes

Fig. 3.6 shows RMSE $\sqrt{N}/\sigma$, where both N and σ are varied and where the scalings of the previous figures are combined. Fig. 3.6a shows that the scaled RMSE of the NLLS estimator is (approximately) constant for a large part of the parameter space ($N > 20$). This is not the case for the CORR estimator, as can be seen in Fig. 3.6b.By comparing Fig. 3.6a and Fig. 3.6b, one can clearly see that the RMSE of the NLLS estimator is much smaller than that of the CORR estimator for any N and σ.

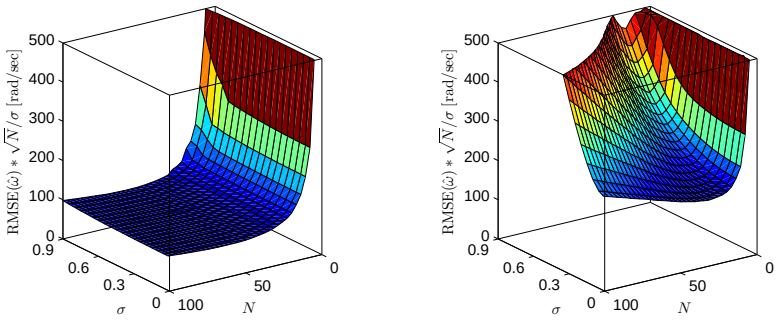
3.4.2 Experimental data

Fig. 3.7 shows the results of the recorded experimental MRI data. Fig. 3.7a shows an SE image of the DTI phantom. This image serves as a suitable basis for comparison of the corrected EPI images. Fig. 3.7b shows the corresponding, original reconstructed EPI image. In this EPI image, a large distortion is visible due to an air bubble located a few slices away, as well as a significant ghosting artifact. Fig. 3.7c shows the field map obtained with the CORR method and Fig. 3.7d shows the field map obtained with the NLLS method. Comparing these images demonstrates that the NLLS method is less sensitive to magnitude differences, as the noise inside the (darker) fiber bundles is clearly lower for the NLLS method. This is more clearly visible in Fig. 3.7e and Fig. 3.7f, which show the standard deviation of the field map of the CORR and NLLS method, respectively. This standard deviation map is computed from 10 complete acquisitions of the reference data. From each of the 10 reference data sets a field map is computed and after subtraction of the median of each field map, the standard deviation of the field map is computed for each pixel. Fig. 3.8 demonstrates the corrected version of the EPI image Fig. 3.7b. Since the field map can only be estimated inside the object, a blue color filter is applied to the signal of the corrected images outside the object, as any signal in this region



(a) (b)

Figure 3.5: *(a) RMSE and (b) RMSE$\sqrt{N}$ of the different methods for the simulated signals with $\omega = 180$ rad/sec. For these results $\sigma = 0.2$, $M = 100\,000$ and the shaded areas represents the 95% confidence regions of the scaled RMSE.*



(a) NLLS (b) CORR

Figure 3.6: *scaled RMSE of the NLLS (a) and CORR (b) estimators as function of σ and N. To more easily compare the results the scaled RMSE is limited to 500. The actual scaled RMSE of (b) is actually (much) higher than 500 for large N and σ. for each point the number of repetitions is $M = 10\,000$.*

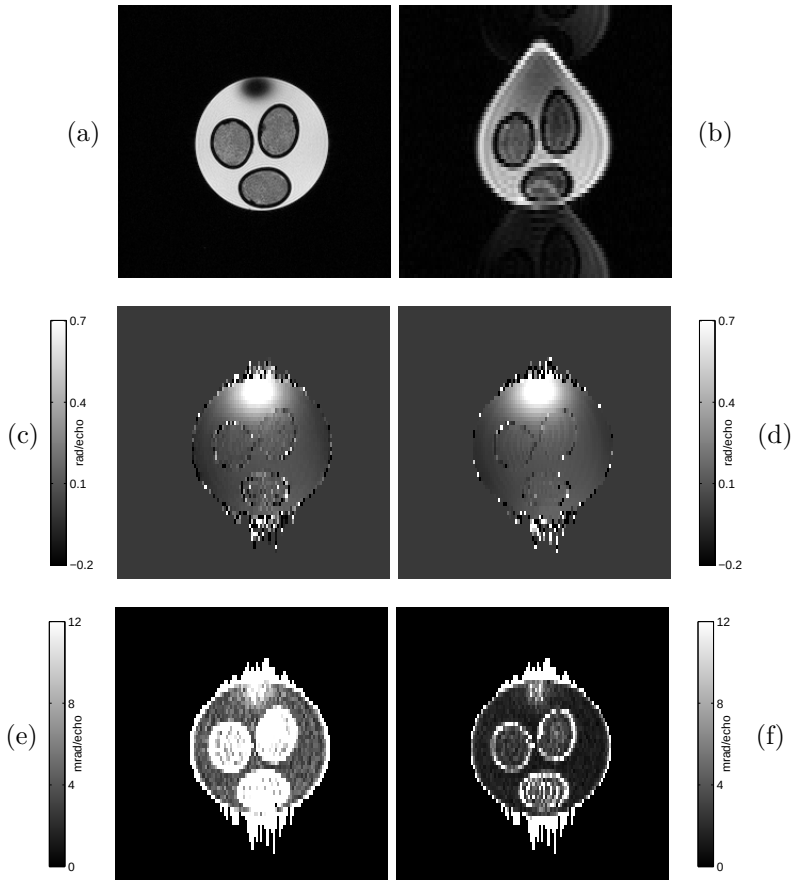


Figure 3.7: *Field map corrections. The images of the DTI phantom object are acquired at the Bio Imaging Lab with a 7 T Bruker Pharmascan small animal MRI system. Figure (a) shows a Spin Echo image of the phantom. Figure (b) shows an EPI image of the same slice. This image shows a large distortion caused by a nearby air bubble. The figures (c) and (d) show a phase map estimated from reference data of this slice for the CORR and the NLLS method, respectively. The background is masked in the field maps, since the field map cannot be estimated when no signal is present. The figures (e) and (f) show the standard deviation of the field map, computed from 10 acquisitions of the field map reference data, for the CORR and NLLS method, respectively.*

is due to off resonance effects of parts of the object and an object mask can be used to remove these spurious signals. Fig. 3.8a and Fig. 3.8b show the results

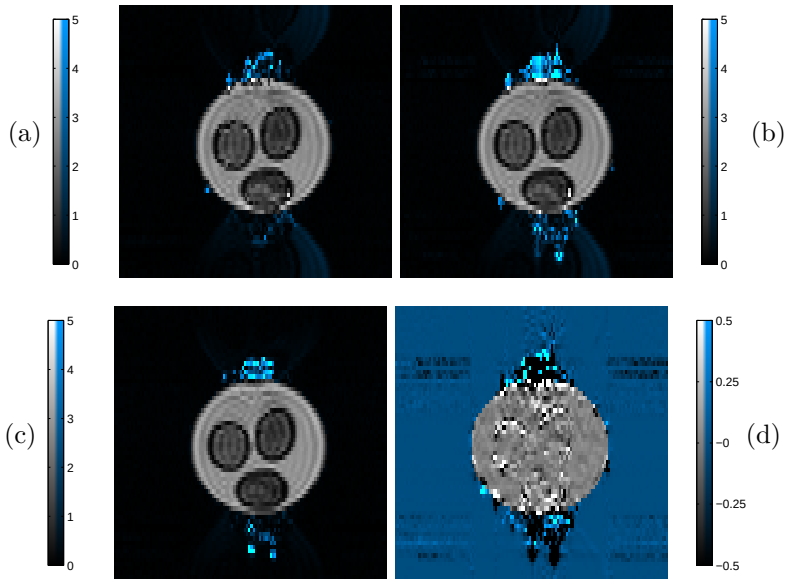


Figure 3.8: *This figure shows the field map correction results. The figures (a) and (b) show the EPI image corrected with the field maps computed with the CORR method and NLLS method, respectively. Outside the object, the field map cannot be estimated, and thus the image cannot be corrected outside the object. Therefore a blue color mask is applied outside the object to indicate that any signal in these regions is not relevant. In figure (c) the ghost of figure (b) is suppressed with the odd-even phase difference estimated by the NLLS method. Figure (d) shows the difference between (a) and (b).*

of the application of the correction scheme to the EPI image with the field map $\hat{\omega}_{\mathrm{COR}}(\boldsymbol{r})$ and $\hat{\omega}_{\mathrm{NLLS}}(\boldsymbol{r})$, respectively. Since no ghost correction is applied in these images, the ghost is clearly visible. In Fig. 3.8c, the even-odd phase difference, which is also estimated by the NLLS estimator, is used to suppress the ghost artifacts still present in Fig. 3.8b. Fig. 3.8d shows the difference between Fig. 3.8a and Fig. 3.8b. As is clearly visible, the largest differences are at the low signal regions around the fiber bundles. In these regions, the correction with $\hat{\omega}_{\mathrm{COR}}(\boldsymbol{r})$ is significantly worse.

In summary, Fig. 3.7 and Fig. 3.8 clearly show the superior performance of the NLLS estimator.

The reference data needed.

The off-resonance map is generally a smooth function of the spatial coordinates. Hence, the spatial resolution with which the field map, and thus the reference data, has to be acquired, may be lower than the resolution of the images that need to be corrected. Moreover, scan time can be reduced by reducing the number of gradient echoes recorded for the reference data. As is shown in Fig. 3.5, the performance of the NLLS estimator improves substantially up to approximately 20 echoes, and for higher N scales with approximately $1/\sqrt{N}$. A further aspect that might be exploited to minimize the reference data scan time is that the reference data does not need to have the same image contrasts as the images to be corrected. As long as the field inhomogeneities are equal, the field map obtained from the reference data can be used for images with different contrasts. Finally, we remark that the substantially improved precision of the proposed NLLS field map estimator compared to the CORR estimators can as well be traded for a faster acquisition of the reference data with reduced SNR (e.g., by reducing TR).

3.5 Conclusions

High speed acquisitions such as EPI, are desirable for techniques like DTI and fMRI. Unfortunately, such acquisitions suffer from serious geometrical distortions, especially at high main magnetic fields. Therefore, correction methods which reduce these distortions are necessary. Such methods estimate the field map, which captures the local magnetic field inhomogeneities. The quality of corrected EPI images depends on the precision and accuracy with which the field map is estimated.

In this chapter, a *nonlinear least squares method* (NLLS) was described to estimate the field map. Compared to a previously proposed estimation method by Smithorst et al. [23], the proposed NLLS was shown to perform substantially better in terms of the root mean squared error of the estimated field map and thus lead to higher quality of the corrected EPI images.

A further benefit of the NLLS estimator is that other parameters of the MR image are simultaneously estimated. These parameters, which include the relaxation and the ghost causing odd/even k-line differences can be used to correct the ghosting and T_2^* blurring.

Acknowledgements

The MR images and reference data presented in this chapter has been acquired at the Bio Imaging lab, University of Antwerp.

3.A Initialization of the optimizations based on the FFT

For convergence to the global optimum of the NLLS estimator, the optimization should be initialized close enough to this optimum. Since a linear phase trend (i.e. frequency) is searched for, good initial values can be found with the Fourier Transform of the signal. When we assume that the odd/even phase jump is small (negligible), the data can be modeled by (compare with Eq. (3.8)):

$$A(j) = e^{\lambda_5 + \mathrm{i}\lambda_2} e^{j(\mathrm{i}\lambda_1 + \lambda_4)}, \tag{3.14}$$

Let

$$A(j) \xrightarrow{\mathcal{F}} \mathcal{A}(\omega) = \frac{e^{\lambda_5 + \mathrm{i}\lambda_2}\left(e^{N(\lambda_4 + \mathrm{i}(\lambda_1 - \omega))} - 1\right)}{e^{\lambda_4 + \mathrm{i}(\lambda_1 - \omega)} - 1} \tag{3.15}$$

be the discrete Fourier transform (z-transform on $e^{\mathrm{i}\omega}$) of $A(j)$, then the position of the maximum absolute value of $\mathcal{A}$ is given by

$$\omega_{\max} = \arg\max_{\omega} |\mathcal{A}(\omega)| \tag{3.16}$$

$$= \lambda_1 \tag{3.17}$$

For normal acquisitions, i.e. when the image should be visible, it can be assumed that the strongest component in A is the exponentially decaying signal. So, the peak at the maximum position $\omega_{\max}$ is from this signal. The parameters $\boldsymbol{\lambda}$ can now be computed from $\mathcal{A}(\omega)$ by

$$\lambda_1 = \omega_{\max} \tag{3.18}$$

$$e^{\lambda_5 + \mathrm{i}\lambda_2} = \mathcal{A}(\omega_{\max}) \frac{1 - e^{\lambda_4}}{1 - e^{N\lambda_4}} \tag{3.19}$$

and λ_4 can be computed from

$$\left.\frac{\partial^2 \log|\mathcal{A}(\omega)|}{\partial \omega^2}\right|_{\omega=\omega_{\max}} = \frac{-e^{\lambda_4}}{(e^{\lambda_4} - 1)^2} + \frac{N^2}{(e^{N\lambda_4} - 1)^2} + \frac{N^2}{e^{N\lambda_4} - 1}. \tag{3.20}$$

For $N \to \infty$ this can be simplified to

$$\left.\frac{\partial^2 \log|\mathcal{A}(\omega)|}{\partial \omega^2}\right|_{\omega=\omega_{\max}, N\to\infty} = \frac{-e^{\lambda_4}}{(e^{\lambda_4} - 1)^2}, \tag{3.21}$$

which can be simplified even more for $|\lambda_4| \ll 1$ to

$$\left.\frac{\partial^2 \log|\mathcal{A}(\omega)|}{\partial \omega^2}\right|_{\omega=\omega_{\max}, N\to\infty, |\lambda_4|\ll 1} = \frac{-1}{\lambda_4^2} + \frac{1}{12}. \tag{3.22}$$

This last expression Eq. (3.22) can easily be solved for λ_4. Since $\frac{-1}{\lambda_4^2} + \frac{1}{12} \geq \frac{-e^{\lambda_4}}{(e^{\lambda_4}-1)^2} \geq \frac{-1}{\lambda_4^2}$ and $\frac{N^2}{(e^{N\lambda_4}-1)^2} + \frac{N^2}{e^{N\lambda_4}-1} > 0$, the true value of

$$|\lambda_4| \leq \sqrt{\frac{-1}{\frac{\partial^2 \log|\mathcal{A}(\omega)|}{\partial\omega^2}\Big|_{\omega=\omega_{\max}}}}. \tag{3.23}$$

As long as the SNR is high enough, the initialization obtained by setting $|\lambda_4|$ to the upper bound of Eq. (3.23) is accurate enough, although the better initialization found by numerically solving Eq. (3.20) might reduce the optimization time enough to be worthwhile. Since there is relaxation (decay of the signal) in the time series, λ_4 is obviously negative. The only parameter not yet initialized by the above method is λ_3, the phase step between the even and odd echoes. Usually this phase step is small, so it can be initialized to zero. Otherwise, the data can be split in an even and odd part and the above initialization can be performed on both parts. The phase step can then be computed from the combination of both parts. In order to numerically compute the second derivative accurately, the discrete Fourier transform $\mathcal{A}(\omega)$ should be smooth enough. This can be achieved by expanding $A(j)$ with zeros before applying the Fast Fourier Transform (FFT). An additional benefit of this expansion is that the signal can be extended to a size for which the FFT is especially fast for any number of echoes N.

Chapter 4

General and Efficient Super-Resolution method for Multi-Slice MRI

A major part of the work described in this chapter has been accepted for presentation at:

the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2010 conference

as: **Dirk H. J. Poot**, Vincent Van Meir, and Jan Sijbers

General and Efficient Super-Resolution method for Multi-Slice MRI

Abstract

In this chapter, a method is developed that reconstructs a high resolution image from an arbitrary set of multi-slice 3D MR images with a high in-plane resolution and a low through-plane resolution. Such images are often recorded to increase the efficiency of the acquisition. With a model of the acquisition of MR images, which is improved compared to previous super-resolution methods for MR images, a large system with linear equations is obtained. With the conjugated gradient method and this linear system, a high resolution image is reconstructed from MR images of an object. In this chapter, a new and efficient method to apply an affine transformation to multi-dimensional images is presented. This method is used to efficiently reconstruction the high resolution image from multi-slice MR images with arbitrary orientations of the slices.

keywords Super-resolution, multi-slice imaging, reconstruction, conjugated gradients, affine transformation, tomographic MRI, multi-dimensional imaging.

4.1 Introduction

This chapter describes how a high resolution isotropic 3D image can be reconstructed from a series of multi-slice 3D MR images, in which the slice thickness is substantially larger than the in plane resolution. The motivation of this work is that MR images, and especially multi-slice images are often acquired with anisotropic voxels. In this case, the slice thickness can be substantially larger than the in-plane resolution. The reason to record MR images with thick slices is the increase in Signal to Noise Ratio (SNR), as the power of the signal emitted by the object scales approximately linearly with the slice thickness. In this chapter, it is shown that several of these multi-slice MR images, recorded with different slice orientations, can be combined into a single, high resolution 3D image with isotropic voxels. A schematic view of such a set of MR images is given in Fig. 4.1. The advantage of multi-slice images, compared to full 3D acquisitions, is that it is possible to interleave the acquisition of slices. That is, while waiting for the relaxation of the magnetization of a slice, (a part of) the k-space of various other slices can be excited and recorded. In general, when the repetition time (TR) is limited by the T_1 decay, it is possible to record multi slice images significantly faster than full 3D images with the same resolution [**?**]. Furthermore, multi slice images might be less influenced by object motion.

Previously, several attempts have been made to improve the resolution of MR images. The methods of Peled et al. [30] and Carmi et al. [31] try to improve the in-plane resolution. The validity of such methods was questioned by Scheffler [32], to which Peled et al. [30] responded. The main criticism, with which we agree, was that the same points in k-space are acquired by MR images shifted in-plane. The in-plane shift introduces a linear phase shift in these k-space samples, so each of the shifted images should contain the same information, except for measurement noise, with no possibility to improve the resolution. Furthermore, the blurring function of [30] was based on only the T_2^* decay, ignoring the blurring due to the finite part of k-space that was sampled, and the up-sampling was performed by linear interpolation, which, as will be demonstrated also in this chapter, introduces aliasing and reduces the magnitude of high spatial frequencies. The method of Carmi et al. [31] models the blurring function by a box function (1 inside the box, 0 outside). This does not properly model the sinc blurring that is due to the sampling inside a box in k-space. Also, the method of [31] is based on exact integer oversampling factors, which limits the applicability. A different method to improve the resolution of MR images is presented by Greenspan et al. [33]. That method improves the resolution in the slice direction, i.e. the direction in

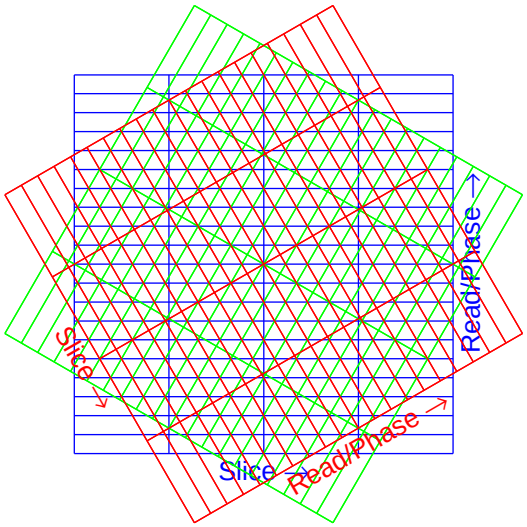


Figure 4.1: *This figure schematically shows three anisotropic MR images acquired with different orientations, rotated around either the phase or read encoding axis.*

which the different slices of a multi slice MR image are recorded. Several multi slice MR images with different positions in the slice direction are combined. This method is not limited to the original resolution as the acquisition in the slice direction is not band limited. However, only MR images in which the slice orientation is identical can be combined. This limitation is removed by the method of Shilling et al. [34], in which multi slice MR images rotated around a common frequency encoding axis are combined. It allows the reconstruction of high resolution slices by iterative projection reconstruction algorithms. In their method they state the projection as a linear system and solve the high resolution image from the set of linear equations by iterative solvers that are also used in Computed Tomography (CT) reconstructions.

In our work, the method of Shilling et al. [34] is extended to allow for any orientation of the slices of the multi-slice MR images, i.e. the images do not need to be rotated around a common frequency encoding axis. Furthermore, the projection via matrix multiplication is reformulated as an affine transformation, followed by a filter operation. This reduces the number of computations substantially. Finally, the method presented in this chapter uses the Conjugated Gradient method to solve the large linear system in a small number of iterations. In this respect, the work presented in this chapter has some relation to the reconstruction of PROPELLOR acquisitions, as described by [35], where

the reconstruction was also performed by the Conjugated Gradient method. Our method, which has the same aim as the method of Greenspan et al. [33], is a super resolution method because the through plane resolution is improved, when a suitable set of differently oriented multi-slice MR images is recorded. This method does not require knowledge about the object. Only, a model of the acquisition system is required. The resolution improvement is obtained by undoing the aliasing and blurring in the slice direction which is due to the acquisition of thick slices.

The structure of this chapter is as follows. In section 4.2, the acquisition model and reconstruction method is explained. Next, in section 4.3, the performance of the method on experimental datasets is shown. In section 4.4, the results are presented and in section 4.5 the conclusions are drawn.

4.2 Methods

4.2.1 Introduction

The acquisition of multiple MR images with the same contrast can be seen as multiple samples from the same object. For each slice of a multi slice MR image, the MR acquisition records a part of the k-space of the object. With the discrete Fourier transform, usually the fast Fourier transform, a projection of the object intensities in each slice is reconstructed on a discrete grid. Since MR acquisitions record a limited part of the k-space, the intensity at a grid point does not exclusively depend on the intensity of the object at the location of that grid point. Essentially, due to the finite part of k-space that is acquired, the MR acquisition effectively applies a low- pass filter to the excited slice of the object before sampling the intensities at the grid nodes. An alternative interpretation is that the intensity value of each voxel of the MR image is obtained by multiplying a properly shifted version of a (3D) sampling function with the object. For the most common multi-slice MR acquisitions, the sampling function can be decomposed in 3 components in orthogonal directions. These 3 components are aligned with the main axis of the grid in the MR image space, defined by the read, phase, and slice encoding directions. In general, when the sampling function is not rotationally invariant and images with different slice positions and/or orientations are recorded, it is possible to reconstruct an image with a higher resolution in at least some directions. In order to explain the method by which the higher resolution image can be obtained, the sampling of the MR images is formalized in the next subsection.

4.2.2 Model of the MRI acquisition

Let $\boldsymbol{x}$ represent the 3D coordinates in the coordinate system of the object o of which the MR images are recorded. Furthermore, let $\boldsymbol{y}$ denote the 3D coordinates in the coordinate system of the j^{th} multi slice MR image S_j. When

both coordinate systems are linked by a coordinate transform T_j, the intensities of S_j are related to those of o by:

$$S_j(\boldsymbol{y}) = \int o(\boldsymbol{x}) w(T_j(\boldsymbol{x}) - \boldsymbol{y}) d\boldsymbol{x} + e_j(\boldsymbol{y}), \tag{4.1}$$

where $e_j(\boldsymbol{y})$ is a noise term that describes the measurement noise. The MR image S_j in Eq. (4.1) is sampled at the integer positions of the MR-image coordinate vector $\boldsymbol{y}$. The sampling function w is (implicitly) defined by the MR image acquisition method. For multi slice acquisition methods that sample a rectangular part of the k-space, this sampling function can be split into 3 functions that are applied in orthogonal directions aligned with the MR- image coordinates, $w(\boldsymbol{y}) = \prod_{i=1}^{3} w_i(y_i)$. Assume, without loss of generality, that the coordinates y_i are ordered 1, 2 and 3 for read encoding, phase encoding, and slice encoding, respectively. Then, due to the rectangular part of k-space that is regularly sampled, w_1 and w_2 are Dirichlet, or periodic sinc, functions. For multi-slice MR images, w_3 depends on the slice selection of the acquisition method. Slice selection is often performed by either a (windowed) sinc or a Gaussian shaped RF pulse, so the sampling in the slice direction w_3 can be modeled by a (smoothed) box or Gaussian function, respectively. Alternatively, it is possible to measure the actual slice excitation profile w_3. For this, in a normal slice acquisition method, only the read encoding gradient direction should be changed to the slice direction and the phase encoding gradient should be set to zero. The recorded radio signal should be Fourier transformed to obtain the slice excitation profile, multiplied by the object integrated over the read and phase encoding directions. When the integrated reference object is (approximately) homogeneous, the excitation profile is obtained.

Note that the transformation $(T_j(\boldsymbol{x}))_2$ is not necessarily just a rotation or even only an affine transformation, but it might also contain other deformations. For example, when the images are recorded with Echo Planar Imaging (EPI) it might also contain displacements in the phase encoding direction due to inhomogeneities of the main magnetic field.

4.2.3 Sampling grid of the object

In Eq. (4.1), a model of the acquisition of MR images from a continuous object is given. However, in order to effectively simulate the acquisition, the object o needs to be discretized. During the MR acquisition, only a part of the k-space is sampled. So, only a low pass filtered version of the object is recorded. Therefore, only the spatial frequencies up to the maximum spatial frequency sampled by the MR images should be representable by the discretized o. This is equivalent to saying that the Nyquist frequency of the grid on which o is discretized should be large enough to contain the part of k-space where the Fourier transform of any w is (significantly) non zero. Note that w, expressed in object coordinates, is different for each different slice orientation. To minimize

the amount of memory used, and probably also the amount of computation time, the number of samples of o should be minimized. Discretizing o by a grid based on a closest sphere packing, such as HCP or FCC, would probably maximize the lowest Nyquist frequency for a fixed number of samples. However, for more convenient post processing, in the remainder of this chapter a regular cubic lattice is used to discretize o.

4.2.4 Discrete model

Let $\boldsymbol{o}$ $(n_o \times 1)$ be a vector containing the object intensities at all grid points $\boldsymbol{x}_m$, $m \in \{1, \ldots, n_o\}$ of the discretized object o. This grid should be sufficiently dense and contain the region in which the object intensity is (might be) non zero. Then, the sampling of a multi slice MR image Eq. (4.1) can be rewritten as a matrix multiplication:

$$\boldsymbol{S}_j = \boldsymbol{X}_j \boldsymbol{o} + \boldsymbol{e}_j \tag{4.2}$$

where $\boldsymbol{S}_j$ $(n_{S_j} \times 1)$ is the j^{th} MR image with a noise term $\boldsymbol{e}_j$ $(n_{S_j} \times 1)$. Both $\boldsymbol{S}_j$ and $\boldsymbol{e}_j$ are sampled at the n_{S_j} nodes of the grid of the j^{th} MR image $\boldsymbol{y}_l$, $l \in \{1, \ldots, n_{S_j}\}$. The elements of the matrix $\boldsymbol{X}_j$ $(n_{S_j} \times n_o)$ are given by :

$$X_j(l, m) = w(T_j(\boldsymbol{x}_m) - \boldsymbol{y}_l) = \prod_{i=1}^{3} w_i(T_j(\boldsymbol{x}_m)_i - (\boldsymbol{y}_l)_i), \tag{4.3}$$

where $T_j(\boldsymbol{x}_m)_i$ and $(\boldsymbol{y}_l)_i$ indicate the i^{th} component of the vectors $T_j(\boldsymbol{x}_m)$ and $\boldsymbol{y}_l$, respectively. The sampling of all N MR images can be combined into a single matrix multiplication

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{o} + \boldsymbol{e}, \tag{4.4}$$

with

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_1 \\ \vdots \\ \boldsymbol{S}_N \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix}, \quad \boldsymbol{e} = \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_N \end{bmatrix}. \tag{4.5}$$

4.2.5 Reconstruction of the object

When the MR images $\boldsymbol{S}$ are acquired, the reconstruction of the object intensities $\boldsymbol{o}$ can be stated as a regularized least squares problem:

$$\hat{\boldsymbol{o}} = \arg\min_{\boldsymbol{o}} |\boldsymbol{X}\boldsymbol{o} - \boldsymbol{S}|_2^2 + |\boldsymbol{K}\boldsymbol{o}|_2^2 \tag{4.6}$$

$$= \arg\min_{\boldsymbol{o}} (\boldsymbol{X}\boldsymbol{o} - \boldsymbol{S})^T (\boldsymbol{X}\boldsymbol{o} - \boldsymbol{S}) + \boldsymbol{o}^T \boldsymbol{K}^T \boldsymbol{K} \boldsymbol{o}, \tag{4.7}$$

where $\boldsymbol{K}$ specifies the regularization term, which will be explained in the next subsection (subsection 4.2.6). In general, the solution of this regularized least squares problem is given by:

$$\hat{\boldsymbol{o}} = (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{X}^T\boldsymbol{S} \tag{4.8}$$

However, for realistic image dimensions, the matrices present in the general solution Eq. (4.8) are too large to actually store, even as sparse matrices (size $\boldsymbol{X} \approx 40M \times 10M$). Furthermore, the solution of the linear system by QR or LU decomposition [**?**] would consume prohibitively many computations for any realistic size of images ($\approx (10M)^3 = 10^{21}$ flops). Therefore, a solution was obtained with the conjugated gradients method [**?**, **?**], which is an efficient iterative method to solve linear systems. The matrix vector multiplications that are required by the conjugated gradient method were evaluated by a function that did not explicitly store $\boldsymbol{X}$. Actually, as will be explained in subsection 4.2.7, the multiplication with $\boldsymbol{X}$ was reformulated as an affine transformation, combined with 1D filter operations.

4.2.6 Regularization

Tikhonov regularization is a standard technique to solve under determined or badly conditioned problems [36]. The current problem is badly conditioned, or even under determined, even though $\sum_{j=1}^{N} n_{S_j}$ usually exceeds n_o. This is due to the high resolution of the grid on which the object intensities are reconstructed. In order to obtain the maximum resolution, all spatial frequencies present in the MR images should fall below the Nyquist frequency of the grid at which the object o is reconstructed. Therefore, this grid will, most likely, also contain (high) spatial frequencies not present in any of the MR images, causing the reconstruction to be badly conditioned or under determined.

The regularization aims to improve the solution by reducing the variance of the solution. However, this immediately introduces a bias into the estimated $\hat{\boldsymbol{o}}$. (Without regularizing, the Least Squares estimator is unbiased.) Therefore, a good choice for the regularization it to choose it such that the Mean Square Error (MSE) is minimized. Since the MSE is the sum of the variance and the squared bias, these are first derived separately.

The variance of $\hat{\boldsymbol{o}}$, assuming normally distributed independent noise with a

standard deviation σ in the MR images, is given by:

$$\begin{aligned}
\text{var}(\hat{\boldsymbol{o}}) &= \mathbb{E}(\hat{\boldsymbol{o}} - \mathbb{E}(\hat{\boldsymbol{o}}))(\hat{\boldsymbol{o}} - \mathbb{E}(\hat{\boldsymbol{o}}))^T && (4.9)\\
&= (\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{X}^T\mathbb{E}(\boldsymbol{e}\boldsymbol{e}^T)\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{K}^T\boldsymbol{K})^{-1} && (4.10)\\
&= \sigma^2\left((\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T) + \boldsymbol{K}^T\boldsymbol{K}\right)^{-1}(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T \\
&\quad (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)\left((\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T) + \boldsymbol{K}^T\boldsymbol{K}\right)^{-1} && (4.11)\\
&= \sigma^2(\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T + \boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T(\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T + \boldsymbol{K}^T\boldsymbol{K})^{-1} \\
&= \sigma^2\boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{V}^T\boldsymbol{K}^T\boldsymbol{K}\boldsymbol{V})^{-1}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{V}^T\boldsymbol{K}^T\boldsymbol{K}\boldsymbol{V})^{-1}\boldsymbol{V}^T, && (4.12)
\end{aligned}$$

where $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is a Singular Value Decomposition (SVD) of $\boldsymbol{X}$, in which $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary ($\boldsymbol{U}^{-1} = \boldsymbol{U}^T$) and $\boldsymbol{\Sigma}$ is non-zero only on the main diagonal. The bias is given by:

$$\begin{aligned}
\text{bias}(\hat{\boldsymbol{o}}) &= \mathbb{E}\{\hat{\boldsymbol{o}}\} - \boldsymbol{o}_0 = \left((\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{K}^T\boldsymbol{K})^{-1}\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{I}\right)\boldsymbol{o}_0 && (4.13)\\
&= \left(\boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{V}^T\boldsymbol{K}^T\boldsymbol{K}\boldsymbol{V})^{-1}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T - \boldsymbol{I}\right)\boldsymbol{o}_0 && (4.14)\\
&= -\boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{V}^T\boldsymbol{K}^T\boldsymbol{K}\boldsymbol{V})^{-1}\boldsymbol{V}^T\boldsymbol{K}^T\boldsymbol{K}\boldsymbol{o}_0, && (4.15)
\end{aligned}$$

where $\boldsymbol{o}_0$ is the true object magnitude. When $\boldsymbol{V}^T\boldsymbol{K}^T\boldsymbol{K}\boldsymbol{V} = \boldsymbol{D}$ and $\check{\boldsymbol{o}}_0 = \boldsymbol{V}^T\boldsymbol{o}_o$, the MSE is given by:

$$\begin{aligned}
\text{MSE}(\hat{\boldsymbol{o}}) &= \text{var}(\hat{\boldsymbol{o}}) + \text{bias}(\hat{\boldsymbol{o}})^2 && (4.16)\\
&= \sigma^2\boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{D})^{-1}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{D})^{-1}\boldsymbol{V}^T + \\
&\quad \boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{D})^{-1}\boldsymbol{D}\check{\boldsymbol{o}}_0\check{\boldsymbol{o}}_0^T\boldsymbol{D}^T(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{D})^{-1}\boldsymbol{V}^T \\
&= \boldsymbol{V}\left(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{D}\right)^{-1}\left(\sigma^2\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{D}\check{\boldsymbol{o}}_0\check{\boldsymbol{o}}_0^T\boldsymbol{D}^T\right)\left(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \boldsymbol{D}\right)^{-1}\boldsymbol{V}^T. && (4.17)
\end{aligned}$$

Since $\boldsymbol{V}$ is a unitary matrix, only the diagonal elements of the part in Eq. (4.17) between $\boldsymbol{V}$ and $\boldsymbol{V}^T$ provide a non zero contribution to the MSE. Therefore, the regularization that minimizes the MSE, assuming $\boldsymbol{D}$ is non-zero only on the main diagonal, is given by:

$$\begin{aligned}
D_{ii} &= \arg\min_{D_{ii}}\left(|\Sigma_{ii}|^2 + D_{ii}\right)^{-2}\left(\sigma^2|\Sigma_{ii}|^2 + D_{ii}^2|\check{o}_{0,i}|^2\right) && (4.18)\\
&= \frac{\sigma^2}{|\check{o}_{0,i}|^2}. && (4.19)
\end{aligned}$$

Note that at first sight this might look as the inverse Signal to Noise Ratio (SNR), but it is not. The scaling of o might differ from the scaling of S of which σ specifies the noise level. There are several reasons why this result cannot be applied directly to specify the regularization. First, in practice,

the true $\boldsymbol{o}_0$ is, by definition, unknown (why would we want to estimate $\boldsymbol{o}$, when it is already known?). However, the regularization is close to optimal when a reasonable estimate of the magnitude of this value (or the SNR) is provided. It can easily be demonstrated that when D_{ii} is a factor 2 from the value that provides the lowest MSE, the MSE might at most increase by a factor 9/8. Therefore, with some general assumptions (e.g. a magnitude of $1/f$ of the spectrum of the image) and an initial reconstruction, obtained with or without a 'default' regularization, a close to optimal value for D_{ii} can be found. Second, as mentioned above, the matrices are large. Actually so large that it is impossible, or at best impractical, to actually compute the SVD for realistic image sizes. It might be possible to study the singular vectors and values of a small problem and deduce a regularization strategy from these. However, in this chapter, we follow a different approach.

In the current problem, the regularization is needed to constrain the high spatial frequencies, as some of these might not be sampled by any of the MR images. Since there is no specific prior knowledge about the high spatial frequencies of the object, the regularization can be used to force the amplitude of the under sampled high frequencies of $\hat{\boldsymbol{o}}$ to zero. This can be achieved by adding the power in the (high) frequencies to the minimization criterium. As especially the high spatial frequencies are under sampled, in this work, the regularization term $\boldsymbol{K}$ computes the second derivatives of the reconstructed $\boldsymbol{o}$:

$$\boldsymbol{o}^T \boldsymbol{K}^T \boldsymbol{K} \boldsymbol{o} = \lambda \left(\left(\frac{\partial^2 \boldsymbol{o}}{\partial {x_1}^2} \right)^2 + \left(\frac{\partial^2 \boldsymbol{o}}{\partial {x_2}^2} \right)^2 + \left(\frac{\partial^2 \boldsymbol{o}}{\partial {x_3}^2} \right)^2 \right) \quad , \tag{4.20}$$

with the simple discrete second derivative, $\left. \frac{\partial^2 o}{\partial {x_i}^2} \right|_{\boldsymbol{x}} = o(\boldsymbol{x}-\boldsymbol{a}_i) - 2o(\boldsymbol{x}) + o(\boldsymbol{x}+\boldsymbol{a}_i)$, where $\boldsymbol{a}_i$ are the base vectors of the grid of o. In the experiments sections, it will be evaluated how closely this regularization matches the one specified by Eq. (4.19). The parameter λ is introduced to scale the regularization to the actual SNR.

4.2.7 Affine transform

It is possible to explicitly evaluate the matrix vector multiplications that are required for the conjugated gradient method by which Eq. (4.8) is solved, by repetitively computing parts of the matrix $\boldsymbol{X}$ and multiplying these with the appropriate parts of the vectors. However, this will require a large amount of computation time. When T_j is an affine transform, or a subset of an affine transform, such as a combination of translation, scaling, and rotation, the acquisition of the MR images can be reformulated as an affine transform of the object o, followed by subsequent filter operations with the three orthogonal sampling functions w_i:

$$o \xrightarrow{T_j} o_j \xrightarrow{*w_1} o_{j,1} \xrightarrow{*w_2} o_{j,2} \xrightarrow{*w_3} S_j \tag{4.21}$$

When implemented efficiently, this allows a substantial increase in performance due to a substantial reduction in the number of operations. However, efficiently applying an affine transform is non trivial. Therefore, several (new) methods by which an affine transform can be applied to an image are studied below. In order to study the properties of these methods, an 'ideal' affine transform is specified, as well as criteria by which the quality of an transform can be evaluated.

An affine coordinate transform between two coordinate systems $\boldsymbol{x}$ and $\boldsymbol{y}$ is a linear transformation specified by

$$\boldsymbol{y} = \boldsymbol{T}\boldsymbol{x} + \boldsymbol{c}, \tag{4.22}$$

where $\boldsymbol{T}$ $n \times n$ specifies the affine transformation matrix in n dimensions, and $\boldsymbol{c}$ $n \times 1$ specifies the translation part of the transformation. Alternatively, the transformation can be specified as

$$\begin{bmatrix} \boldsymbol{y} \\ 1 \end{bmatrix} = \boldsymbol{T}_f \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix}, \tag{4.23}$$

with

$$\boldsymbol{T}_f = \begin{bmatrix} \boldsymbol{T} & \boldsymbol{c} \\ \boldsymbol{0} & 1 \end{bmatrix}. \tag{4.24}$$

The affine transformation of a continuous image o is specified by

$$\tilde{o}(\boldsymbol{x}) = o(\boldsymbol{T}\boldsymbol{x} + \boldsymbol{c}), \tag{4.25}$$

where $\tilde{o}$ is the continuous transformed image. This transform of a continuous image specifies the ideal transform. However, as only a finite set of samples of an object can be stored, this ideal transform can only be approximated. When the image is band limited, i.e. it has a finite range of spatial frequencies, is periodic, and is sampled on a sufficiently dense regular grid, then all information of the continuous image is present in this finite sample. However, the grid points at which the transformed image $\tilde{o}$ is requested are, in general, not equal to the grid points on which the source image o is given. Hence, some kind of 'interpolation' is needed. Also note that after a general affine transform, the original periodicity of the image will not be maintained. Thus some distortions are unavoidable. However, in practice, these distortions can be limited to the border region, both of the spatial and frequency domain.

In order to properly evaluate the errors introduced by a transformation method, it is required to study both the spatial and the frequency domain. Fortunately, when an affine transformation is applied to an image, the frequency domain is distorted by an affine transformation and a phase factor. Every vector $\boldsymbol{f}$ $(1 \times n)$ in the frequency domain of o, is transformed to a vector $\boldsymbol{f}'$ in the frequency domain of $\tilde{o}$ by :

$$\boldsymbol{f}' = \boldsymbol{f}\boldsymbol{T}^{-1}, \tag{4.26}$$

and the translation, specified by $\boldsymbol{c}$, introduces a phase shift , which is linear in $\boldsymbol{f}$, in the frequency domain. Thus

$$\tilde{\mathcal{O}}(\boldsymbol{f}) = \mathcal{O}(\boldsymbol{f}\boldsymbol{T}^{-1})e^{i\boldsymbol{f}\boldsymbol{c}}, \tag{4.27}$$

where $\mathcal{O} = \mathcal{F}(o)$, $\tilde{\mathcal{O}} = \mathcal{F}(\tilde{o})$. This transformation allows the proper study of errors introduced into the frequency domain when applying an affine transformation. For the study of the transformation methods we assume that the domain of spatial frequencies at which the image $\mathcal{O}$ has a non zero magnitude is given by a convex set of frequency vectors $\boldsymbol{f}$ with $|\boldsymbol{f}|_\infty \leq f_n$, where $f_n = .5$ is the Nyquist frequency in Hz. The spatial extent of o is assumed to be given by a convex set of spatial vectors $\boldsymbol{x}$. With these assumptions, already an approximation is introduced, since the extent of an image cannot be finite in both the spatial and frequency domain. However, when the image, potentially after proper extrapolation in the spatial domain, smoothly reaches zero at the borders, the errors introduced by this approximation are very (arbitrarily) small.

The transformation method by which an image is transformed might differ in computational cost and/or accuracy of the transform. In this section the computational cost is approximated by the average number of image samples that is processed for a voxel of the destination image. Even though this might not capture all performance influencing factors, it will allow accurate ordering of the methods as long as the performance is not limited by memory bandwidth, which usually is true for efficient implementations of the advanced methods. The accuracy of a transformation method is a combination of several different aspects:

Geometrical distortions: Are the object intensities displaced by the amount specified by the transform?
This obviously is an important criterium. However, all methods that will be considered introduce no erroneous geometrical distortions, so this property does not discriminate between the methods studied below.

Spectral distortions: Are the (magnitudes of the) spatial frequencies of the source image distorted in the output image?
The simple methods, e.g. the linear interpolator, do distort the spatial frequencies, see Fig. **??**. However, it is relatively easy to prevent or repair these spectral distortions. The distortion d is evaluated with

$$d = \int_{\boldsymbol{f}\in\boldsymbol{\Omega}} (h(\boldsymbol{f}) - 1)^2 \, d\boldsymbol{f}, \tag{4.28}$$

where $\boldsymbol{\Omega}$ is the region of spatial frequencies that are present in the source image, i.e. all frequencies below the Nyquist frequency, and can be represented in the destination image, and h is the transfer function of the spatial frequencies from source to destination image.

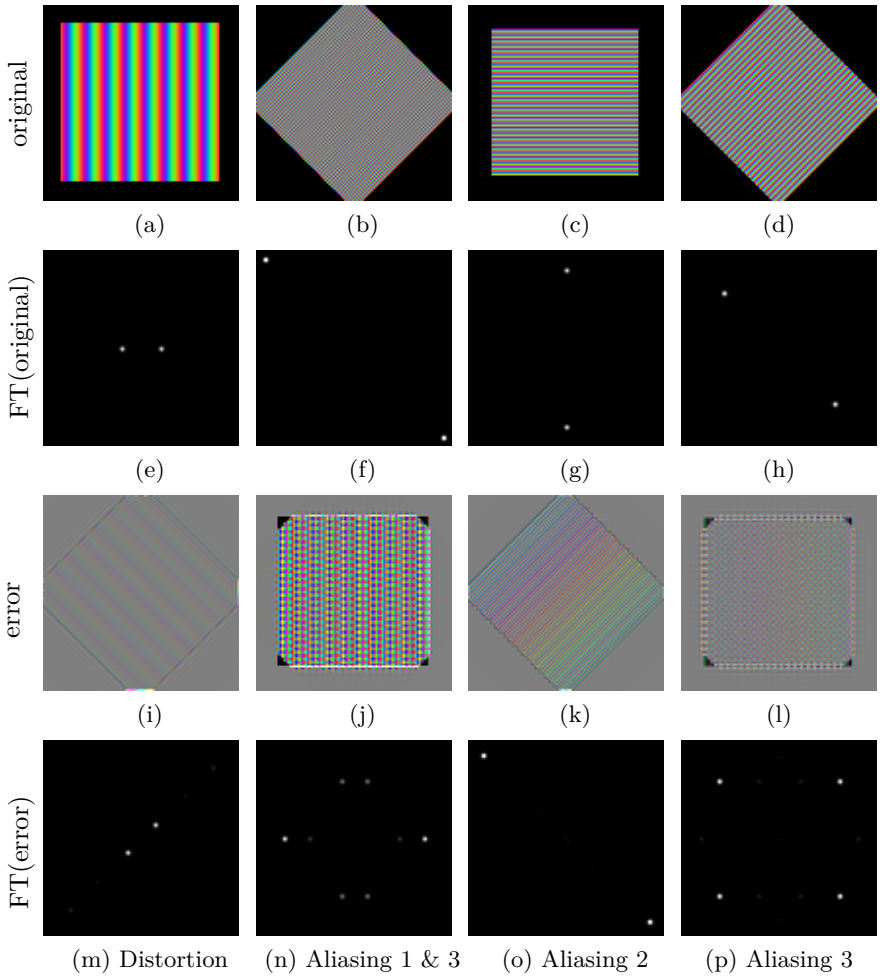


Figure 4.2: *This figure demonstrates spectral distortion and aliasing that can be introduced by transformation methods. (a,b,c,d): Original images, a box with a single spatial frequency, with 120 degree phase difference between the red green and blue (RGB) channels. (e,f,g,h): Fourier transform of the images. The center is the zero spatial frequency, the edges are at the Nyquist frequency. (i,j,k,l): Difference in the spatial domain between the original image rotated over 45 degrees using linear interpolation and a 'perfectly' rotated version. To only focus on aliasing, the spectral distortion introduced by linear interpolation was corrected, except for the images displaying the distortion (left most column) (m,n,o,p): Fourier transform of the error images, i.e. the erroneous energy in the different spatial frequencies. (a,e,i,m): Distortion due to linear interpolation, which is obvious even for this rather low spatial frequency (20% of the Nyquist frequency). (b,f,j,n): Aliasing 1 & 3. In (n) the brightest spots: aliasing 1, the lower intensity spots: aliasing 3. (c,g,k,o): Aliasing 2. (d,h,l,p): Aliasing 3.*

Aliasing: Is there some power of spatial frequencies of the source image that is aliased to different (thus: wrong) spatial frequencies in the destination image?
This is an often overlooked, but important property of transformation methods. It is difficult to avoid aliasing and, in general, it cannot be corrected after the transform has been applied. In practice, it cannot always be completely avoided, but by carefully designing the transformation, the amount aliasing can be minimized. The aliasing in multidimensional ($n > 1$) images might arise from three subtly different sources, which are described below and also visualized in Fig. 4.2. In order to interpret these different forms of aliasing, the transformation procedure is abstractly decomposed in three steps: First, interpolation of the discrete image o to a continuous image. Second, affine transformation of this continuous image. Third, sampling of $\tilde{o}$ from this transformed continuous image.

Aliasing1 The first source of aliasing, Aliasing1, are the spatial frequencies that are below the Nyquist frequency in the source image o ($f_{Ny,s}$), but are transformed to above the Nyquist frequency in the destination image $\tilde{o}$ ($f_{Ny,d}$).

Aliasing2 The second source of aliasing, Aliasing2, are the spatial frequencies that, due to aliasing introduced by the interpolation of the sampled image $\boldsymbol{o}$, are above $f_{Ny,s}$, but are transformed to below $f_{Ny,d}$.

Aliasing3 The third source of aliasing, Aliasing3, are the spatial frequencies that are above $f_{Ny,s}$ in the continuous source image and above $f_{Ny,d}$ in the continuous destination image. When $\tilde{o}$ is sampled from this continuous destination image, these frequencies alias to below $f_{Ny,d}$.

In this subsection several different methods by which the affine transform can be applied are considered. These methods are briefly listed here, and will be explained in more detail below:

LIN **LIN**ear interpolation (LIN) computes the samples of the transformed image $\tilde{o}$ by linear interpolating on the samples of the original image o.

RES Higher order direct **RES**ampling methods (RES), which uses higher order, i.e. longer than the length 2 of LIN, sampling functions. This includes cubic interpolation and (windowed) sinc interpolation of the source image.

URD Subsequently **U**p-sampling, **R**esampling with linear interpolation, and **D**own-sampling (URD). As will be explained below, this reduces some of the problems of LIN and RES.

Table 4.2: *Computational cost per voxel of the output image, distortion, and aliasing when rotating a 2D image by 45 degree. An indication of the computational cost of a 3D rotation is indicated as well. The filterlength specifies the number of lobes of the filter, so the number of samples used in a single multiplication of the filter with the image is (filterlength $*$ Nyquist frequency /cutoff frequency). This ensures that the transition bandwidth in the final image is the same for each filter, and thus the total distortion is minimized. For the FRF method, the filterlength is the number of samples used at the resampling stage.*

	# comp 2D	# comp 3D	filterlength	distortion	Aliasing
LIN	4.0	8.0	2	0.317	0.441
RES	121.0	1331.0	11	0.010	0.171
URD	228.5	940.7	10	0.016	0.011
FRF	124.5	650.7	6	-.-	-.-
SSH	122.9	241.7	16	0.002	0.002

FRF **F**ourier transform with zero expand, **R**esample in the frequency domain, and inverse **F**ourier transform and select the result image. This is an alternative which also reduces some of the problems of LIN and RES.

SSH **S**plit the affine transform in a series of **SH**ear operations, so each of the $2n$ steps can be performed with a (good) 1D interpolation.

In the remaining of this section, it is assumed that the grid points at which the image is or should be known are located at the integer positions of either the source or destination image. So o is known at the integer positions of $\boldsymbol{x}$ and $\tilde{o}$ is requested at the integer positions of $\boldsymbol{y}$. In the subsections below, the effects of the methods are demonstrated with a 45 degree rotation. The figures display the frequency domain response. The approximate computational complexity, as well as the distortion and aliasing, of the 45 degree rotation of each method is presented in Table 4.2. Note that in this 2D example, the SSH method was set up to have a computational cost (almost) equal to the RES method, which leads to distortions and aliasing well below that of both the RES and the even more expensive URD methods.

LIN and RES

The linear interpolation (LIN) is a special case of the more general higher order interpolation methods (RES). Therefore, they will be explained simultaneously in this subsection. These methods are closely related to the specification of the sampling of the MR images given in Eq. (4.1). For these methods, implicitly, the discrete source image, which can be represented as scaled Dirac delta functions at each grid node, is convolved with a continuous interpolation function. The

destination image is then sampled from this continuous image:

$$\tilde{o}(\boldsymbol{y}) = \sum_{\boldsymbol{x}} o(\boldsymbol{x}) w(\boldsymbol{T}\boldsymbol{y} + \boldsymbol{c} - \boldsymbol{x}). \tag{4.29}$$

For LIN, the interpolation function in each dimension is given by

$$w_{\mathrm{LIN}}(x) = \begin{cases} 0 & x < -1 \\ x+1 & -1 \leq x < 0 \\ 1-x & 0 \leq x < 1 \\ 0 & 1 \leq x \end{cases} \tag{4.30}$$

Note that $w_{\mathrm{LIN}}(x)$ is nonzero for $-1 < x < 1$, which is an interval of length 2. The interpolation functions of the other methods of RES will most often have a larger interval in which they are non zero. For example, a sinc function with a raised cosine window of length l_w is given by:

$$w_{\mathrm{RES}}(x) = \begin{cases} 0 & -l_w/2 \geq x \\ \frac{1}{2}\left(1 + \cos(\frac{2\pi x}{l_w})\right) \mathrm{sinc}(x) & -l_w/2 < x < l_w/2 \\ 0 & l_w/2 \leq x \end{cases} \tag{4.31}$$

For an interpolation function with a non zero interval of length l_w, the computational cost in the number of samples for each element of an n dimensional image $\tilde{o}$ is

$$c_{\mathrm{RES}} = l_w^n. \tag{4.32}$$

The transfer function h_{LIN} is the n dimensional Fourier transform of w:

$$h_{\mathrm{RES}}(\boldsymbol{f}) = \mathcal{F}(w(\boldsymbol{x})), \tag{4.33}$$

which for the linear interpolation is given by

$$h_{\mathrm{LIN}}(\boldsymbol{f}) = \prod_{i=1}^{n} \mathrm{sinc}(f_i)^2. \tag{4.34}$$

In order to prevent large distortions and a substantial amount of Aliasing2 and Aliasing3, the filter lengths need to be large, which might require prohibitively many computations for $n \geq 2$. Even with long filters, a substantial amount of Aliasing1 might still be present. See for example Fig. 4.3 for the aliasing that is present with a 45 degree rotation in 2D. This aliasing is inherent to any direct interpolation method. Finally, note that each sample of $\tilde{o}$ is sampled individually from o. Thus, this resampling method can also be applied to non affine transformations.

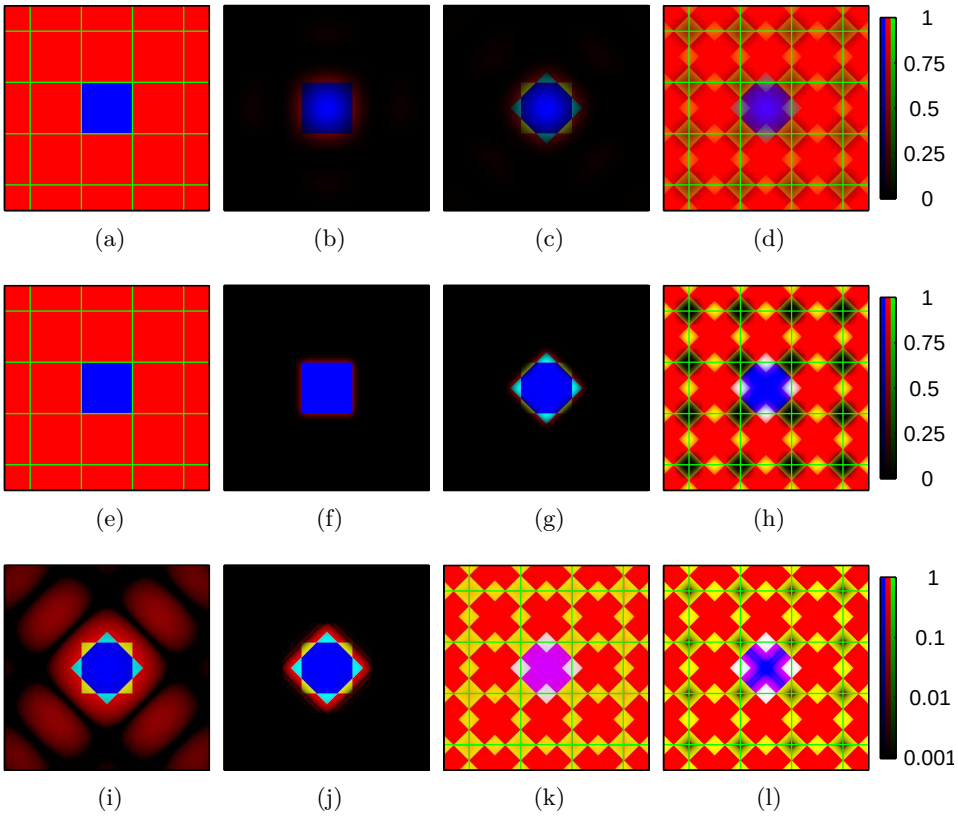


Figure 4.3: *Frequency domain distortions and aliasing when rotating over 45 degrees and resampling by linear interpolation (a-d) or by the RES interpolator with $l_w = 10$ (e-h). Note that the continuous images of which the frequency domain is shown exist only implicitly. (a),(e) is the original image, periodic in the frequency domain. The central mode is colored blue, the non central modes, which will contribute to Aliasing3, are colored red. (b), (f) The continuous image after interpolation by the 2D interpolator. (c), (g) The continuous image after rotating. (d), (h) After sampling the continuous rotated image. To show the aliasing more clearly, (i), (j), (k), and (l) show the logarithm of (c), (g), (d), and (h), respectively. The green lines indicate the borders of the elementary frequency cells. Only the blue part in these images is not distorted due to aliasing. The intensity variations in the blue part are due to the spectral distortions. The different aliasing contributions are colored: Aliasing1: Cyan, Aliasing2: Yellow/Olive, Aliasing3: Purple (& Red)*

URD

The Aliasing1 that could not be avoided by the RES method, can be removed by first up-sampling, then resampling, then down-sampling the image. The up-sampling with a low pass filter creates an area in the frequency domain where (almost) no energy is present. It can be ensured that at the resampling step, only the area in the frequency domain without energy does contribute to Aliasing1, but this might require sampling on a grid that is more dense than required for $\tilde{o}$. Therefore, the final stage in the URD method is down-sampling with a lowpass filter. Schematically, the URD method is given by:

$$o \overset{\text{upsampling}}{\longrightarrow} o_a \overset{\text{resampling}}{\longrightarrow} o_b \overset{\text{downsampling}}{\longrightarrow} \tilde{o} \tag{4.35}$$

See Fig. 4.4 for an example where an image is rotated over 45 degrees by the URD method.

The up and down-sampling that is present in this method can be applied in every dimension separately, like the filter operations in Eq. (4.21). In this method the actual transformation is still direct sampling as described in subsection 4.2.7. Hence, the computational cost strongly depends on the length of the interpolation function w. However, there are three reasons why, for the same quality of the result image, the length of the interpolation function can be substantially shorter than in the RES method. First, the distortion caused by w can be pre-compensated in the up-sampling stage. Second, interpolators typically introduce the largest errors close to the Nyquist frequency. After up-sampling, the frequency domain is non empty only close to the zero frequency, thereby reducing the Aliasing3. Thirdly, the main contribution to Aliasing3 is caused by the first of the periodic aliases in the frequency domain, see Fig. 4.4e. By carefully choosing the up and down-sampling factors, the aliasing caused by these modes can be moved outside the area in the frequency domain that remains after the low-pass filtering for the down-sampling step, reducing Aliasing3.

The computational cost of the URD method is

$$c_{\text{URD}} = l_u * \sum_{i=1}^{n} u^i + d^n * l_w^n + l_d * \sum_{i=1}^{n} d^{i-1}, \tag{4.36}$$

where l_u is the length of the up-sampling filter, u the up-sampling factor, l_d the length of the down-sampling filter, d the down-sampling factor, where it was assumed that the up and down-sampling factors and filter lengths are equal along every axis. This is not required, but simplifies the design and computation of the cost. Furthermore, it has been assumed that the up and down-sampling is implemented as explicit convolution with finite length low-pass filters. Alternatively, the up and down-sampling operations might be implemented by zero filling or extraction in the frequency domain. With the Fast Fourier Transform (FFT), the computational cost might be lower when

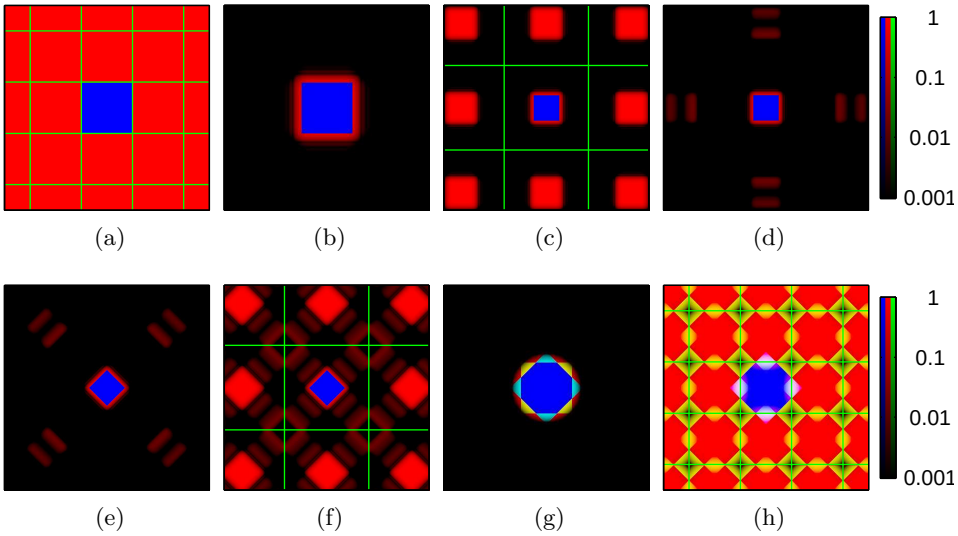


Figure 4.4: *Logarithm of the frequency domain of the individual steps when rotating over 45 degrees with the URD method.(a) The original image, which is periodic in the frequency domain. The central mode is colored blue, the non central modes, which will contribute to Aliasing3, are colored red. (b) Continuous image after convolving with a lowpass filter (with $l_u = 10$). (c) Up-sampled image, (factor 3.3). (d) Continuous image, obtained by 2D linear interpolation of the up sampled image. (e) The continuous image after rotating. (f) Upsampled destination image, obtained by sampling the rotated image on grid with 3.3 times higher density than the result image. (g) is after low pass filtering ($l_d = 33$) the up-sampled result image. (h) Result image, obtained by sampling with the final grid density. The subfigure (h) can be directly compared to Fig. 4.3 (k) and (l). The blue part is non aliased and non constant intensity in the blue part indicates spectral distortions. The green lines indicate the borders of the elementary frequency cells. The other colors indicate aliasing. The different aliasing contributions are colored: Aliasing1: Cyan, Aliasing2: Yellow/Olive, Aliasing3: Purple (& Red)*

l_u or l_d are large. Depending image size and implementation, typically around $l > 40$.

Note that the up-sampled images that are created in the URD method might require a large amount of memory to store. With a 3D image and an up-sampling factor of 3.3, which was used for the 45 degree rotation presented in Fig. 4.4, the temporary up-sampled image requires $3.3^3 \approx 36$ times as much memory as the original image. The memory consumption might be reduced by carefully ordering and interleaving the up-sampling, re-sampling and down-sampling computations. It is quite easily possible to reduce the memory consumption to slightly more than what is needed for the image after up-sampling or before down sampling, whichever is smaller.

Finally, note that also this URD method can be applied to non affine transformations, as long as it is possible to obtain the deformations at the nodes of the up-sampled grid.

FRF

A substantially different way of affinely transforming an image is by applying the resampling in the spatial-frequency domain. So the transformation is applied with Eq. (4.27):

$$\boldsymbol{o} \xrightarrow{\mathcal{F}} \mathcal{O} \overset{\text{resampling}}{\longrightarrow} \tilde{\mathcal{O}} \xrightarrow{\mathcal{F}^{-1}} \tilde{\boldsymbol{o}} \tag{4.37}$$

With this method, the spatial and frequency domain are effectively switched compared to the previous methods. Therefore, the Aliasing3 that was present in the previous methods will manifest itself as fold over artefacts in the final image $\tilde{\boldsymbol{o}}$. By expanding o with zeros and sampling $\mathcal{F}(\tilde{\boldsymbol{o}})$ on a sufficiently dense grid, these fold over artefacts can be moved outside of the final image. The use of proper sampling filters should ensure that the non central modes do not alias to inside the image. For this, the intensity of these non central modes should be reduced to well below the noise level of the image. The resampling filters should be designed specifically for each expansion factor, and together with the expansion factor they influence the intensity of the fold over artefacts. For example, with a filter length of 8 samples, and a 2 fold zero expansion in every dimension, the relative intensity of the fold over artefacts can be as low as 10^{-4}. The main advantage of this method is that it avoids all aliasing of spatial frequencies, at the expense of fold over artefacts.

Note that the resampling in the frequency domain that is required in this method is very closely related to the resampling/gridding that is performed for a non uniform FFT [37, 38].

The computational cost is the cost of 2 n dimensional FFT's and a resampling which has a cost l_w^n. The length of the filters will usually be quite large in order to avoid fold over artefacts of the non central modes. Furthermore, as the image needs to be extended with zeros to reduce l_w, the memory consumption might be too high.

SSH

The affine transform $\boldsymbol{T}_f$ can also be applied as a series of 1D shear transforms. There exists a (non unique) set of transformations $\tilde{\boldsymbol{T}}_j$, different from the identity matrix only in row d_j, such that $\boldsymbol{T}_f = \prod_{j=1}^{2n} \tilde{\boldsymbol{T}}_j$. Each partial transform $\tilde{\boldsymbol{T}}_j$ is a shear transform and modifies the image only along one of the $d_j{}^{\text{th}}$ main axis of the image. Therefore, each individual sub transform can be efficiently applied by a 1D low pass filter. A general affine transform in n dimensions can be split in $2n$ shear operations, which combined have arbitrarily low aliasing and arbitrarily low distortion. This only requires the assumption that filters (arbitrarily) close to ideal low-pass filters can be constructed. Schematically:

$$\boldsymbol{o} \xrightarrow{\tilde{\boldsymbol{T}}_1} \boldsymbol{o}_a \xrightarrow{\tilde{\boldsymbol{T}}_2} \dots \xrightarrow{\tilde{\boldsymbol{T}}_{2n}} \tilde{\boldsymbol{o}} \tag{4.38}$$

In order to avoid aliasing and distortions, the set of frequencies that are below $f_{Ny,s}$ and which are transformed to below $f_{Ny,d}$, $\boldsymbol{\Omega}$, need to be below the Nyquist frequency after every intermediate transformation. Since elements of $\boldsymbol{\Omega}$ might be close to the Nyquist frequency, this in general requires that the frequency vectors should not be modified in a specific dimension, unless a large enough empty area in the frequency domain has been created by up-sampling in that dimension.

With the following procedure, the transform $\boldsymbol{T}_f$ is decomposed in a set of shear transforms that satisfy these criteria:

- Initialize $\tilde{\boldsymbol{T}}_j$ for $j \in \{1 \dots 2n\}$ to the $n+1 \times\ n+1$ identity matrix.
- By computing the total cost of applying the transformations defined in the following steps, search for the optimal permutation p of the numbers $\{1, \dots, n\}$.
- Initialize the initial remaining transform to the inverse transformation: $\boldsymbol{R}_0 = \boldsymbol{T}_f^{-1}$
- for $j = 1 \dots n$:

$$(\tilde{\boldsymbol{T}}_j)_{p_j, p_{1\dots j}} = \left(\left((\boldsymbol{R}_{j-1})_{p_{1\dots j}, p_{1\dots j}} \right)^{-1} \right)_{j, 1\dots j} u_j \tag{4.39}$$

$$\boldsymbol{R}_j = \tilde{\boldsymbol{T}}_j \boldsymbol{R}_{j-1}, \tag{4.40}$$

  where $(\boldsymbol{A})_{a,b}$ selects the rows a and the columns b, of the matrix $\boldsymbol{A}$.

- for $j = n+1 \dots 2n$

$$(\tilde{\boldsymbol{T}}_j)_{p_{j-n}, 1\dots n+1} = \left(\boldsymbol{R}_{j-1}^{-1} \right)_{p_{j-n}, 1\dots n+1} \tag{4.41}$$

$$\boldsymbol{R}_j = \tilde{\boldsymbol{T}}_j \boldsymbol{R}_{j-1}. \tag{4.42}$$

- Find the lowest up-sampling factors u_j, such that at every step the maximum spatial frequency that should be transfered to the final image in the dimension d_j is below the lowest aliased frequency in that dimension. Changing the permutation order p might influence these minimal required up-sampling factors. The cutoff frequency of the 1D filter of step j should be set to the average of the maximum spatial frequency that should be transfered to the final image in the dimension d_j and the lowest aliased frequency in that dimension.

See Fig. 4.5 for this method applied to a 45 degree rotation.

We believe that splitting the general affine transformation with this procedure leads to the lowest number of shear transformations by which all aliasing can be avoided (with ideal low pass filters). As can easily be verified, the transformations $1 \dots j-1$ and $j+n+1 \dots 2n$, $j \leq n$ do not alter the spatial frequencies in the dimension p_j. For some transformations, such as those close to a 90 degree rotation, the amount of shear of the partial transforms, and therefore the up-sampling factors and number of computations, might be reduced considerably by first permuting the source image, or, equivalently, by permuting the destination image after the set of shear transformations has been applied. Since the number of dimensions will always be small (1, 2, 3, and possibly, 4), all permutations can be tested to search for the combination that minimizes the number of computations. However, note that the number of operations in this exhaustive search is of the order of $\mathcal{O}((n!)^2 n^4)$, so for large n ($n > 5$) the exhaustive search is not practical.

The number of computations in this method is approximately

$$c_{\text{SSH}} = l_u * \sum_{i=1}^{n} \prod_{j=1}^{i} u_j + l_d * \sum_{i=1}^{n} \prod_{j=i+1}^{n} d_j, \tag{4.43}$$

which depends on the up and down sampling factors. These strongly depend on the actual transform, but are typically much lower than of the URD method. Comparing this with Eq. (4.36) immediately shows the lower computational cost of the SSH method. Furthermore, note that subsequent filter operations, such as those specified in Eq. (4.21), can be incorporated in the down-sampling filters by convolving the low-pass filter that is applied for the down-sampling in dimension d_j with w_{d_j}.

Note that this transformation method is specifically designed for affine transformations, so non affine transformations can in general not be evaluated by this method.

Comparison of transformation methods

As last part of this section about affine transformations, the different transformation methods are compared. First, Fig. 4.6 shows the aliasing and distortion

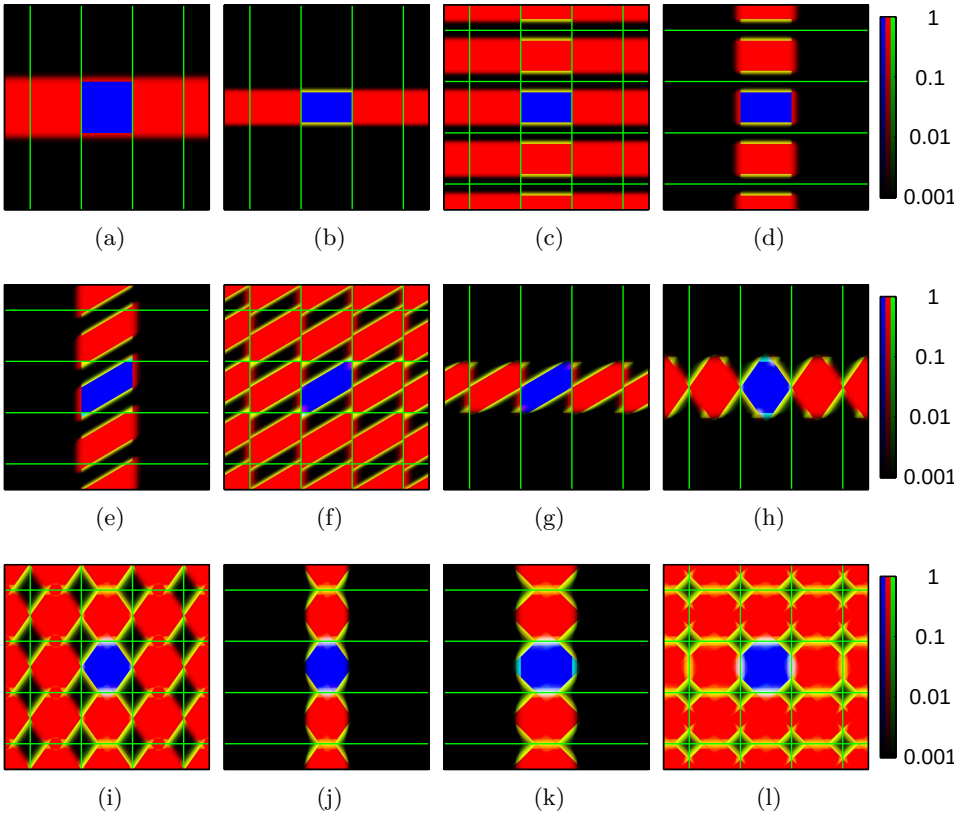
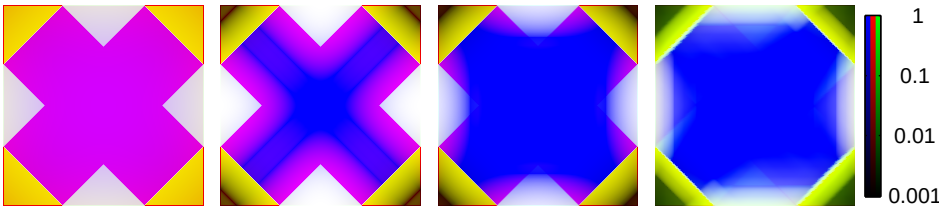


Figure 4.5: *Logarithm of the frequency domain of the individual steps when rotating over 45 degrees with the SSH method. The original image, which is equal to Fig. 4.4(a) is not shown. (a) Image continuous in the vertical direction after convolving the original image with a lowpass filter in the vertical direction. (b) Increased resolution in the vertical direction. (c) Sampled in the vertical direction (d) low pass filtered in the horizontal direction (e) Shear applied in the horizontal direction, note that this causes a vertical shear in the frequency domain. (f) After sampling in horizontal direction (g) After low pass in vertical direction. (h) After shear in vertical direction. Note that this vertical shear causes a horizontal shear in the frequency domain. (i) After sampling in the vertical direction. (j) After low pass filtering in horizontal direction. (k) After down sampling in horizontal direction. (l) The final image after sampling in the horizontal direction. This subfigure can be directly compared to Fig. 4.3(k) and (l) and Fig. 4.4(h). The blue part is non aliased. The green lines indicate the borders of the elementary frequency cells. The other colors indicate aliasing. The different aliasing contributions are colored: Aliasing1: Cyan, Aliasing2: Yellow/Olive, Aliasing3: Purple (& Red)*



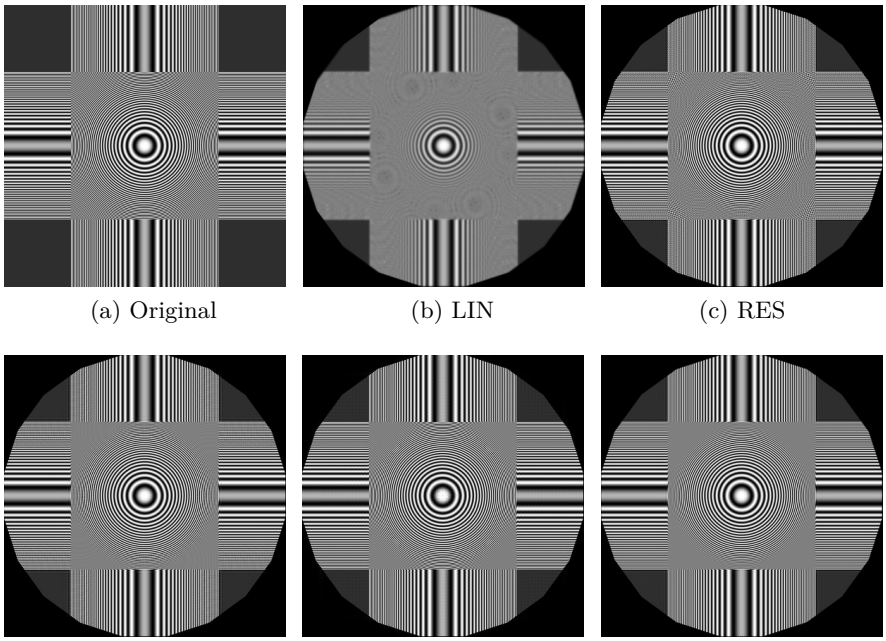
(a) LIN (b) RES (c) URD (d) SSH

Figure 4.6: *Logarithm of the frequency response of the different methods. In this figure only the unit frequency cell of the final image is shown. For explanation of the colors, see caption of Fig. 4.5*

present in the unit frequency cell of the 45 degree rotated images. This distortion and aliasing is due to the transformation methods. Fig. 4.6 is a combination of the central frequency cell of the final transformed images presented in the previous Figures 4.3, 4.4, and 4.5. As is clearly visible, the SSH method has the lowest distortion and least aliasing. Next, Fig. 4.7 shows a test image after 10 rotations of 36 degrees, rotated by all methods that have been explained. Note that the LIN method strongly attenuates the high frequencies. With the RES method with windowed sinc interpolation, the aliasing is clearly visible in the corners of the central square. These corners should have a constant intensity, as the spatial frequencies in the original image cannot be represented in some of the intermediate rotated images. The FRF method has the lowest distortions of the spatial frequencies, but a substantial amount of Gibbs ringing shows up, substantially more than in the URD and SSH methods. Of these two methods the SSH method has the lowest amount of frequency distortion, almost as low as the FRF method. Note that with our implementation, the computational time required by the SSH method was essentially equal to the time required by the LIN method, which used the MATLAB linear interpolation routine `tformarray`. The other methods required substantially more computational time, although a substantial part of this might be due to lower effort put in the optimization of this code.

4.2.8 Why not reconstruct in k-space?

The method described in this chapter reconstructs the high resolution isotropic image in image space directly. Alternatively, one might consider the reconstruction of the k-space of the high resolution image. Since both the acquisition of the MR image, as well as the Fourier transform, are linear operators, an equation equivalent to Eq. (4.3) can be given in k-space. However, in the slice direction, the maximum frequency in which $\mathcal{F}w_3$ is non zero is substantially larger than the Nyquist frequency. This implies that the periodicity in the slice



(a) Original (b) LIN (c) RES

(d) URD (e) FRF (f) SSH

Figure 4.7: *Original test image (a), and the images obtained after rotating the test image in 10 steps over a total of 360 degrees with the different affine transformation methods (b)-(f). When viewing the PDF version of this thesis, please zoom to at least 200% to avoid interference with the screen resolution.*

direction of the k-space of a specific MR image, which is due to the distance between the acquired slices, needs to be taken into account. Without any detailed study, we think that this would increase the complexity of a k-space version of Eq. (4.3), which would most likely increase the amount of computations that are required for the reconstruction. An advantage of reconstruction in k-space might be that it could potentially be easier to include and correct distortions due to inhomogeneities of the magnetic field.

4.3 Experiments

Several datasets of a bird were recorded at the Bio Imaging lab, University of Antwerp, with a Bruker small animal scanner. The resulting MR images were $192 \times 192 \times 32$ with voxel dimensions $0.125mm \times 0.125mm \times 0.75mm$. MR images in $N = 36$ different orientations were recorded and the reconstructed volume spanned the whole head in a volume of $21mm \times 21mm \times 22mm$ with

isotropic voxel dimensions $0.1mm \times 0.1mm \times 0.1mm$.

The alignment of the images as obtained by the scanner was found to be insufficient for accurate reconstruction. Therefore, as preprocessing step, the MR images were properly aligned by computing the required translations from the projections of the MR images to the object space,

$$\boldsymbol{o}_j = \boldsymbol{X}_j^T \boldsymbol{S}_j. \tag{4.44}$$

Since the images are assumed to have the same contrast, the optimal translation between two images was computed by a mean square difference measure,

$$\hat{\boldsymbol{\Delta}}_{j,k} = \underset{\boldsymbol{\Delta}_{j,k} \in \mathcal{R}^3}{\arg\min} \sum_{\boldsymbol{x}} \left(o_j(\boldsymbol{x}) - o_k(\boldsymbol{x} + \boldsymbol{\Delta}_{j,k}) \right)^2. \tag{4.45}$$

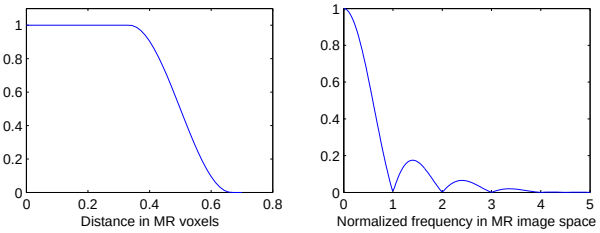
The position of each image j was adjusted by $1/N \sum_{k=1}^{N} \hat{\boldsymbol{\Delta}}_{j,k}$.

After the alignment of the images, the conjugated gradient method was used to approximately solve Eq. (4.8). The matrix vector products that are part of this method are evaluated by the SSH affine transformation method. As the SSH method assumes band limited signals, the low-pass sampling functions w_1 and w_2, which are due to the finite part of k-space that is sampled in the read and phase encoding directions, do not need to be explicitly included. The only non ideal low pass sampling function that is present in the MR acquisition is the slice selection function w_3. All MR images are acquired with the same Hermite slice selection RF pulse. The spatial slice selection function w_3 is the Fourier transform of this slice selection RF pulse and thus is approximated by a smoothed box function. In our reconstruction experiments, it was observed that the exact shape of w_3 did not strongly influence the reconstructed images, so a relatively simple smoothed box function was used:

$$w_3(y_3) \quad = \quad \begin{cases} 1 & |y_3| \leq \frac{1}{3} \\ \frac{1}{2} - \frac{1}{2} \sin\left(3\pi(|y_3| - .5)\right) & \frac{1}{3} < |y_3| < \frac{2}{3} \\ 0 & \frac{2}{3} \leq |y_3| \end{cases} \tag{4.46}$$

See Fig. 4.8 for a graphical representation of this slice selection function in MR image coordinates, both in spatial as well as in frequency domain.

The strength of the regularization is controlled with the variable λ in Eq. (4.20). In our reconstructions λ was set to 5. In order to interpret this value it is important to know the magnitude scaling between the MR images and the reconstructed image. In our implementation, w was normalized in the MR image space, causing the magnitude of the reconstructed object to be reduced by the ratio of the voxel volumes, which was $(1.25^2 * 7.50 = 11.72)$. When we assume that by choosing $\boldsymbol{K}$ to be the second derivative, $\boldsymbol{D}$ happens to be (approximately) diagonal and the singular vectors (approximately) select specific spatial frequencies, $D_{ii} = 3 * 6\lambda$ for i corresponding to singular vectors that select the in-plane Nyquist frequency of an MR image. Then, the MSE



(a) spatial domain (b) frequency domain

Figure 4.8: *The slice selection sampling function w_3, in spatial (a) and frequency (b) domain.*

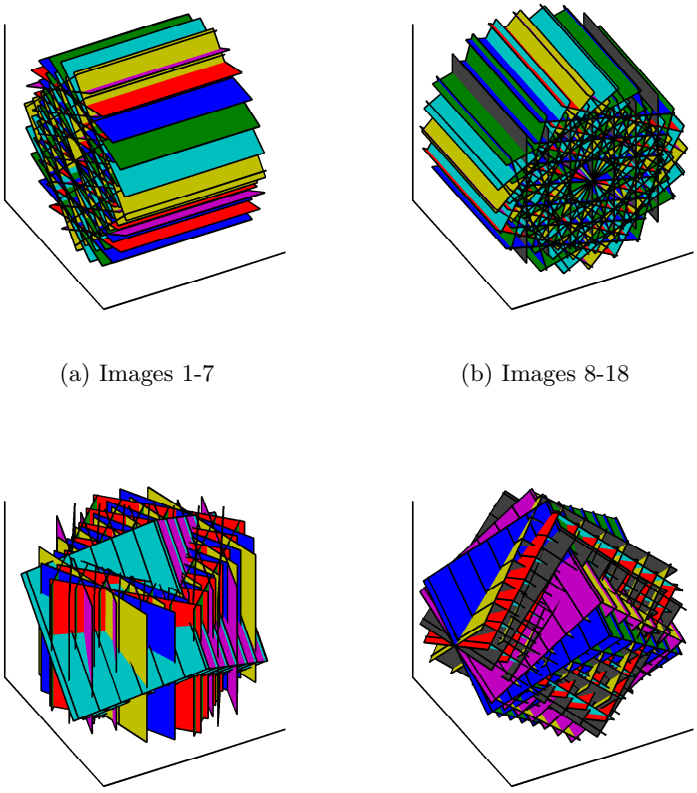
is optimal for $|\check{o}_{0,i}|^2 = \frac{\sigma^2}{D_{ii}} = \frac{\sigma^2}{3*6\lambda}$, which, optimizes the MSE for a SNR of $\sqrt{(11.7187^2/(5*6*3))} = 1.2$ at the Nyquist frequency of MR images. We investigated whether this very rough approximation has any connection to reality. For this, the matrix $\boldsymbol{X}$ was explicitly computed for a problem in which the object space was reduced to $15 \times 15 \times 15$ voxels and MR image space was reduced as well. The geometry of the MR images was the same as in the experiment. First, it was observed that most singular vectors indeed approximately contain spatial frequencies of a specific magnitude. Next, it was observed that $D_{ii} \approx 3*6\lambda$ for the i corresponding to the singular vectors that select the Nyquist frequencies in the MR images. Also, for these i, $\sum_{k\neq i} |D_{ki}| \approx 4D_{ii}$ and $\sqrt{\sum_{k\neq i} |D_{ki}|.^2} \approx \frac{1}{4}D_{ii}$, indicating that $\boldsymbol{D}$ is approximately diagonal. Thus, we can be reasonably confident about the approximation.

See Fig. 4.9 for the orientations of the 36 slices that were recorded. Note that the first 3 groups are acquisitions rotated around one of the slice encoding axis (except for 1 MR image that whose orientation was specified incorrectly) and the last group is rotated around a different axis.

With our implementation of the SSH affine transformation method, each iteration of the conjugated gradient method, with all 36 MR images, took approximately 5m:35s of CPU time on one core of a Intel Core 2 Quad CPU @ 3.0 GHz, with 8GB of RAM.

4.4 Results

One of the original MR images is shown in Fig. 4.11. This 3D MR image is displayed in the read-phase, read-slice, and phase-slice directions, which, for this image, coincide with the coronal, sagittal and transversal directions. Note the substantially lower resolution of the MR image in the slice direction. Fig. 4.12 shows the high resolution image reconstructed from all 36 MR images.



(a) Images 1-7

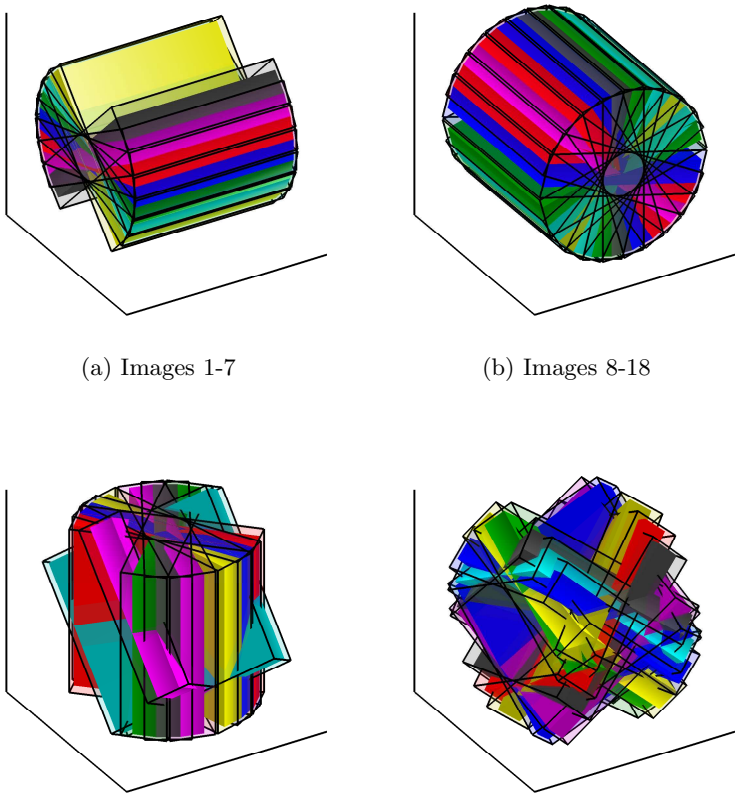
(b) Images 8-18

(c) Images 19-26

(d) Images 27-36

Figure 4.9: *This figure displays every 5th slice of each of the 36 multi slice MR images, all planes of a multi slice MR image have the same color. The 36 images are separated in 4 groups to prevent an overly cluttered view.*

Fig. 4.12a, 4.12b and 4.12c show intersections of the reconstructed object in the same planes as Fig. 4.11. Note the substantial improvement of the resolution in the slice direction and the reduction of the noise. Furthermore, note the improved detail in the coronal views, which is due to the reduced blurring in the slice direction. Fig. 4.13 shows reconstructions with reduced sets of images. The high resolution image was reconstructed with each of the groups of Fig. 4.9. Note that especially group (a) has artefacts, which are due to the substantial part of k-space that is not sampled by this group of images, which is also visible in Fig. 4.10. Furthermore, note that the noise level is slightly higher for these



(a) Images 1-7

(b) Images 8-18

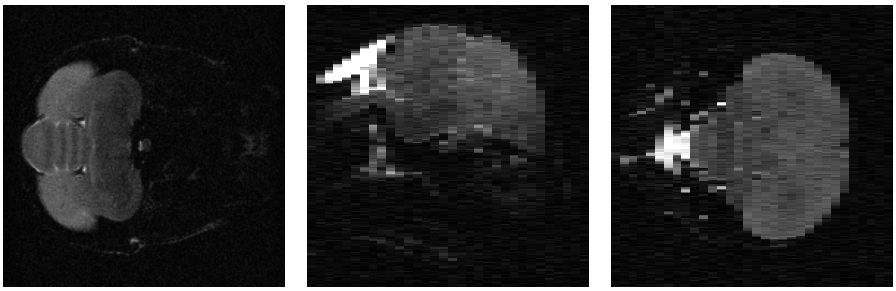
(c) Images 19-26

(d) Images 27-36

Figure 4.10: *This figure schematically displays the part of k-space that each of the 36 multi slice MR images sample. Each k-space box that is sampled by a multi slice MR image has the same color as the slices of that image in Fig. 4.9. The solid inner box contains the spatial frequencies at which the frequency response of the slice selection is larger than .8, the outer, highly transparent box contains the spatial frequencies whose magnitude is higher than the maximum sidelobe.*

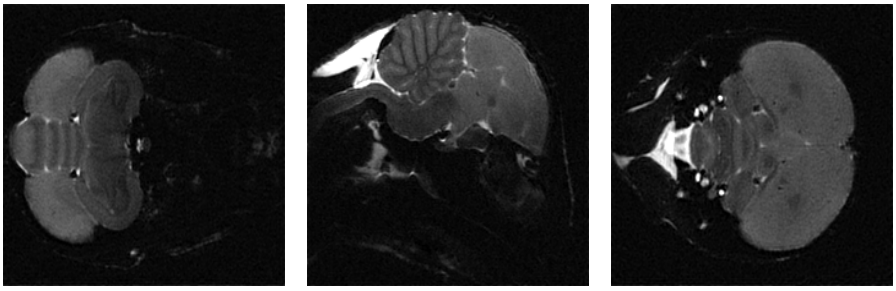
images.

Fig. 4.14a shows the progression of the conjugated gradient method. The blue curve, $|\boldsymbol{o}_{100} - \boldsymbol{o}_i|$, shows the progression of the norm of the difference between an intermediate reconstruction $\boldsymbol{o}_i$ at iteration i and the final reconstruction after 100 iterations. The green curve, $|\boldsymbol{o}_i - \boldsymbol{o}_{i-1}|$, shows the norm of the update in each iteration. Since each update of the conjugated gradient



(a) coronal (b) sagittal (c) transversal

Figure 4.11: *One of the 36 original MR images. Intersections along the 3 main directions are displayed.*



(a) coronal (b) sagittal (c) transversal

Figure 4.12: *The high resolution reconstructed image. Intersections along the 3 main directions are displayed.*

method is orthogonal with respect to the previous updates, $\sum_{j=i}^{99} |\boldsymbol{o}_j - \boldsymbol{o}_{j+1}|^2 = |\boldsymbol{o}_{100} - \boldsymbol{o}_i|^2$. The red curve, $|\boldsymbol{r}_i|$, shows the residue norm at each iteration i: $\boldsymbol{r}_i = \boldsymbol{X}^T(\boldsymbol{S} - \boldsymbol{X}\boldsymbol{o}_i) - \lambda \boldsymbol{K}\boldsymbol{o}_i$, which is zero in the true solution of the regularized least squares solution. As is clearly visible, the magnitude of the update of $\boldsymbol{o}$ quickly reduces and after 15 iterations it is approximately 0.001 of the magnitude of $\boldsymbol{o}$. After these 15 iterations, the difference to the final image is approximately 0.01 of the magnitude of $\boldsymbol{o}$. Even though the update of $\boldsymbol{o}$ is still non-zero, only a limited number of iterations is needed before the error of the solution due to the limited number of iterations is small compared to the noise that is transfered to the reconstructed image. Note that this number of iterations needed for convergence is substantially lower than the approximately 1000 iterations required by the methods in [34]. This higher rate of convergence of the Conjugated Gradient method directly translates to faster reconstruction times. Fig. 4.14b shows the relative updates, $|\boldsymbol{o}_i - \boldsymbol{o}_{i-1}|/|\boldsymbol{o}_i|$,

of the reconstructions with the different groups of images. This image demonstrates that the convergence speed is reduced when a lower number of images is used for a reconstruction.

4.5 Conclusion

In this chapter, a method was developed by which a high resolution isotropic image can be reconstructed from a set of anisotropic multi-slice MR images, recorded with different slice orientations. In contrast to previous reconstruction methods, this new method does not constrain the slice orientations. The reconstruction method uses an improved model of the MR acquisition, but does not require any prior knowledge about the imaged object. The high resolution image of the object is accurately reconstructed by the conjugated gradient method in a small number of iterations, substantially less than previous methods. The experiments show that the quality of the reconstructed isotropic image is substantially better, both in resolution and SNR, than any of the original MR images.

In order to perform the matrix multiplications that are present in the Conjugated Gradient method with a reasonable amount of computational power, this matrix multiplication was reformulated as affine transformation combined with a filter operation. A fast method to affinely transform images while avoiding aliasing as well as distortions in the frequency domain was developed. This method splits a general affine transformation in a set of shear operations, each of which can be efficiently applied by a 1D low pass interpolation filter.

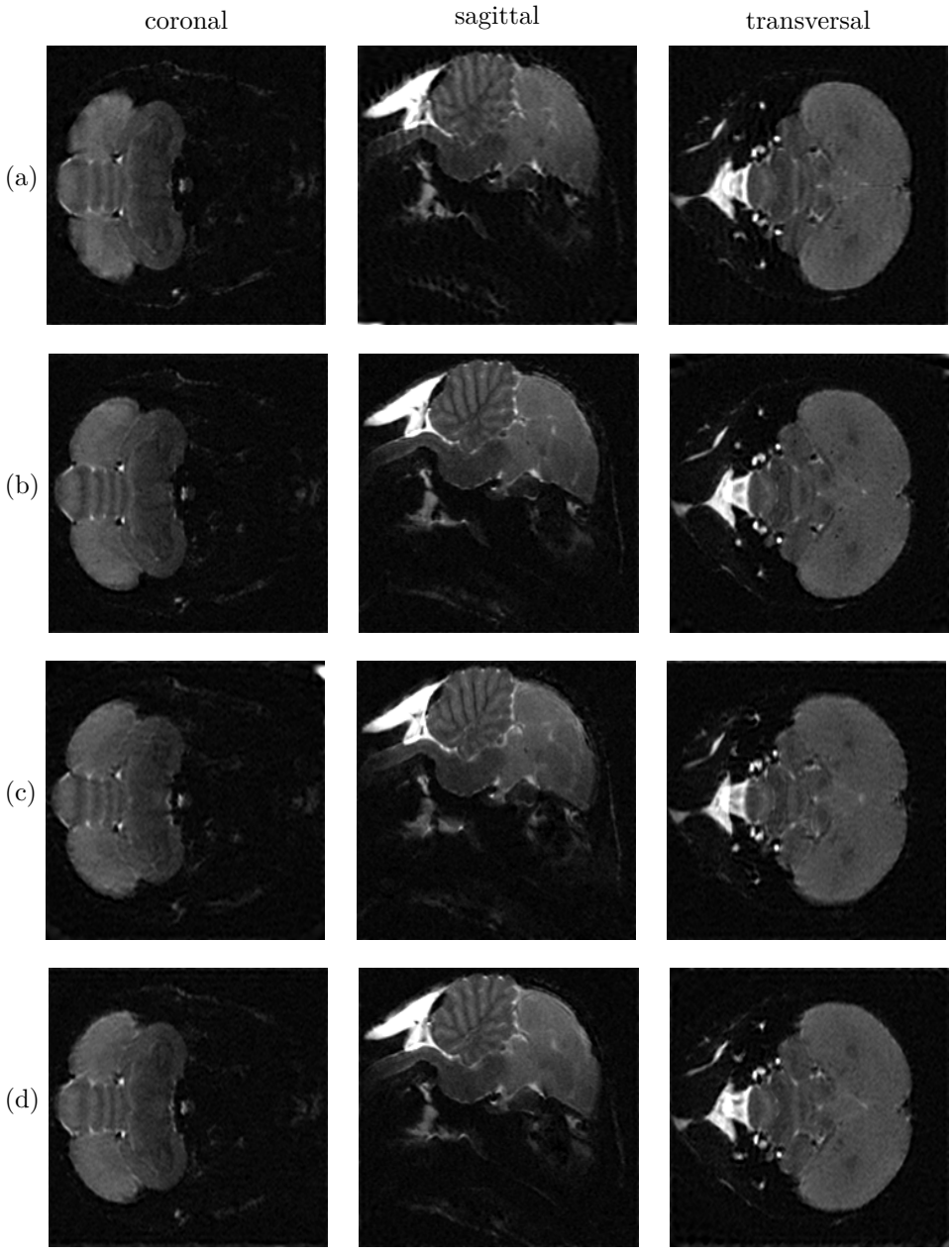
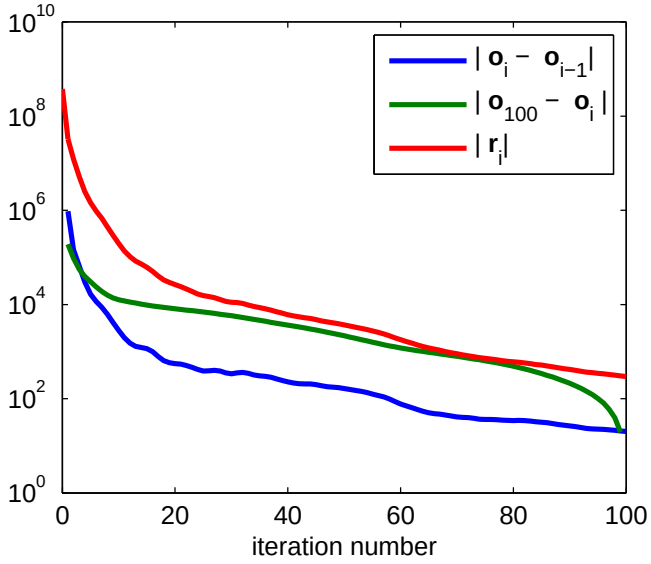
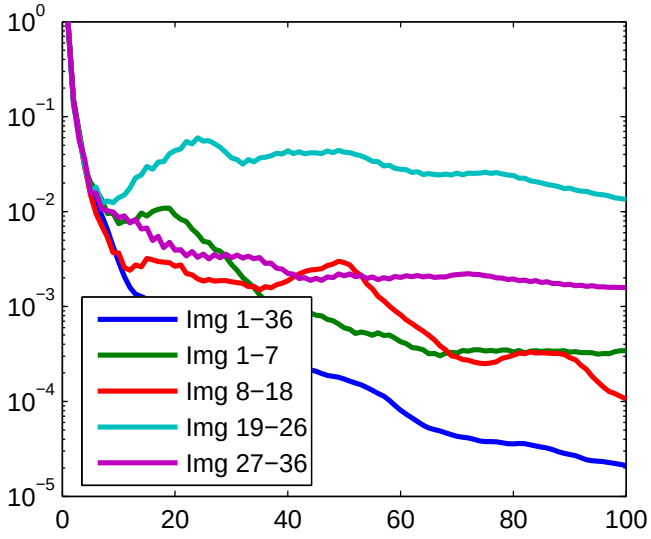


Figure 4.13: *The high resolution reconstructed image, reconstructed with the 4 groups of images also selected in Fig. 4.9 and Fig. 4.10. Intersections along the 3 main directions are displayed.*



(a)



(b)

Figure 4.14: *(a) Norm of update, difference with final, and residue during the conjugated gradient iterations, when all 36 images are used for the reconstruction. (b) Relative update at each iteration for the different groups of images.*

Chapter 5

Automatic estimation of the noise variance of MR images

The work described in this chapter has been published as:

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

in: Physics in Medicine and Biology,
Volume 52, February 2007, Pages 1335-1348

Abstract

Estimation of the noise variance of a magnetic resonance (MR) image is important for various post-processing tasks. In the literature, various methods for noise variance estimation from MR images are available, most of them however requiring user interaction and/or multiple (perfectly aligned) images. In this chapter, we focus on automatic histogram-based noise variance estimation techniques. Previously described methods are reviewed and a new method based on the maximum likelihood (ML) principle is presented. Using Monte Carlo simulation experiments as well as experimental MR data sets, the noise variance estimation methods are compared in terms of the root mean-squared error (RMSE). The results show that the newly proposed method is superior in terms of the RMSE.

5.1 Introduction

The noise variance in magnetic resonance (MR) images has always been an important parameter to account for when processing and analyzing magnetic resonance imaging (MRI) data. Algorithms for noise reduction, segmentation, clustering, restoration, and registration highly depend on the noise variance [39–42]. Also, many applications that employ statistical analysis techniques, such as functional MRI or voxel based morphometry, often base their conclusions on assumptions about the underlying noise characteristics [43–45]. Finally, knowledge of the noise variance is useful in the quality assessment of the MR imaging system itself, for example to test the noise characteristics of the receiver coil or the preamplifier [46].

In the past, many techniques have been proposed to estimate the image noise variance. These can be subdivided into two classes:

multiple images In the past, noise variance estimation methods were developed based on two acquisitions of the same image. A standard procedure was developed by Sano in which the noise variance was estimated by subtracting two acquisitions of the same object and calculating the standard deviation of the resulting pixel values [47, 48]. Multiple acquisition methods are relatively insensitive to structured noise such as ghosting, ringing, and DC artifacts [49, 50]. However, strict requirements are the perfect geometrical alignment of the images and temporal stationarity of the imaging process.

single image The image noise variance can also be estimated from a single magnitude image. A common approach is to estimate the noise variance from a large, manually selected, uniform signal region or non-signal (i.e., noise only) region [10, 12, 14, 51, 52]. Manual interaction however clearly suffers from inter and intra operator variability. An additional problem is that the size of the selected (homogeneous) regions should be sufficiently large to obtain a precise estimate of the noise variance. Moreover, background data may suffer from systematic intensity variations due to streaking or ghosting artifacts.

Often, magnitude MR images contain a large number of background data. Hence, the noise variance can as well be estimated from the background mode of the image histogram. Automatic noise variance estimation have been designed from the knowledge that this background mode can be represented by a Rayleigh distribution [53, 54]. In this chapter, these procedures are reviewed and a new method is presented.

In this chapter, we describe a new method to estimated the image noise variance from the background mode of the image histogram. Our initial motivation to search for a new method was that existing methods that exploit

this background mode for the same purpose, seemed to be based on heuristic arguments, leaving significant space for finding an improved method.

In Section 5.2.1, we will describe previously reported procedures to estimate the noise variance from the background mode of the image histogram. Then, in Section 5.2.2, we will present a new noise variance estimation method based on maximum likelihood (ML) estimation from a partial histogram. Subsequently, in Section 5.3 and 5.4, the performance of the described noise variance estimation procedures in terms of precision and accuracy are evaluated and discussed, respectively, for simulated as well as experimental data sets. Finally, in Section 5.5, conclusions are drawn.

5.2 Methods

Recall that in Section 2.5 was proven that the distribution of the non-signal background areas is Rayleigh distributed.

5.2.1 Previously reported, histogram-based noise variance estimation methods

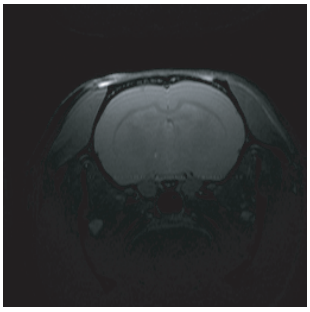
Magnitude MR images generally contain a large number of background data points. Hence, the histogram of such images often shows a background mode that is clearly distinguishable from the signal contributions in the histogram. As an example, in Fig. 5.1, three coronal spin-echo MR images of a mouse brain are shown along with the corresponding histogram. The images, of size 256×256, were acquired on a 7 Tesla SMIS MR imaging system, using a field of view of 30 mm in both directions. Fig. 5.1a shows a proton density weighted image (TE=20 ms, TR=3000 ms), Fig. 5.1c a T_2 weighted image (TE=60 ms, TR=3000 ms), and Fig. 5.1e a T_1-weighted image (TE=20 ms, TR=300 ms). As can be observed from Fig. 5.1b, Fig. 5.1d, and Fig. 5.1f, a background mode can easily be observed.

To estimate the noise variance from the image histogram background mode, automatic and robust noise variance estimation methods have been reported that exploit this background mode along with the knowledge that the noise-only contribution represents a Rayleigh distribution [53–55]. In this section, these methods are reviewed. Next, in subsection 5.2.2, a new method is described based on ML estimation.

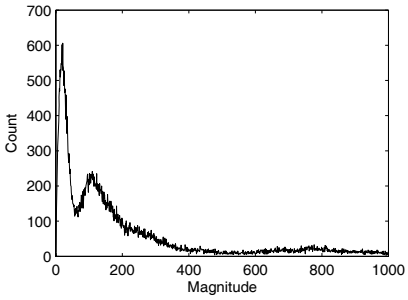
Maximum of the background mode of the histogram

From the Rayleigh PDF, given in Eq. (2.9), the noise variance can be estimated by searching for the value of m for which the Rayleigh PDF attains a maximum [55]:

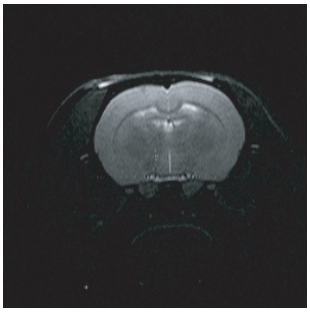
$$\frac{\partial p}{\partial m} = 0 \qquad \Leftrightarrow \qquad 1 - \frac{m^2}{\sigma^2} = 0 \quad . \tag{5.1}$$



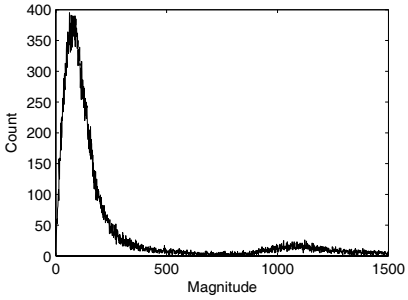
(a) Proton density weighted



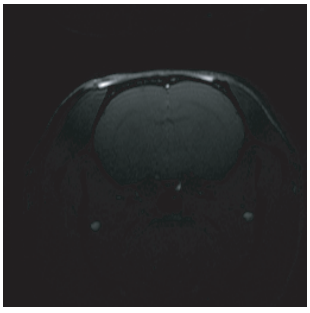
(b) Histogram of Fig. 5.1a



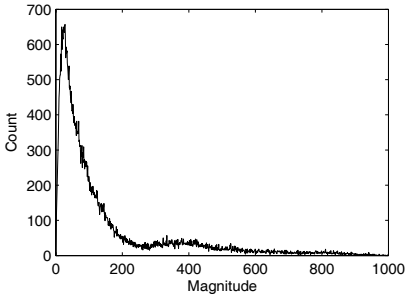
(c) T_2 weighted



(d) Histogram of Fig. 5.1c



(e) T_1 weighted



(f) Histogram of Fig. 5.1e

Figure 5.1: *2D coronal MR image and corresponding histogram of a mouse brain: (a-b) proton density-, (c-d) T_2-, (e-f) T_1-weighted image.*

From this, it is clear that an estimate of the noise standard deviation is simply given by:

$$\widehat{\sigma} = m_{max} \quad . \tag{5.2}$$

In practice, $m_{\max}$ can easily be found by searching for the magnitude value at which the background mode in the histogram attains a maximum. Since the background mode is always located on the left side of the histogram, finding this maximum is trivial.

Brummer

In the work of Brummer et al. [53], a noise variance estimation method is presented in which the Rayleigh distribution is fitted to a partial histogram using least squares estimation:

$$\widehat{N}_{\mathrm{Br}}, \widehat{\sigma}_{\mathrm{Br}} = \arg \max_{N_{\mathrm{Br}}, \sigma_{\mathrm{Br}}} \sum_{f=0}^{f_c} \left(h(f) - N_{\mathrm{Br}} \frac{f}{\sigma_{\mathrm{Br}}^2} e^{-(f^2/2\sigma_{\mathrm{Br}}^2)} \right)^2 \quad , \tag{5.3}$$

where N_{Br} is the amplitude and σ_{Br} the width of the Rayleigh distribution that is fitted to the histogram h. The cutoff f_c is defined as

$$f_c = 2\sigma_{Br,0} \quad , \tag{5.4}$$

where $\sigma_{Br,0}$ is an initial estimate of the noise level. Brummer's method specifies that the position of the first local maximum of the low-pass-filtered grey-value histogram is to be used as the initial estimate. In our implementation of Brummer's method, we used Chang's estimate (see subsection 5.2.1) as an initial value.

Chang's noise variance estimation method

In order to improve robustness of the noise variance estimation method described in the subsections 5.2.1 and 5.2.1, Chang et al. proposed a procedure to smooth the histogram prior to estimation [54]. Thereby, a Gaussian smoothing kernel

$$\kappa(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, \tag{5.5}$$

was used. The smoothing width h was set to

$$h = 1.06\sigma_0 n^{1/5} \tag{5.6}$$

in which σ_0 is the sample standard deviation and n the sample size. The smoothed histogram at signal level x is given by

$$\widehat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} \kappa \left(\frac{x - x_i}{h} \right), \tag{5.7}$$

where $\{x_i\}$ is the image intensity data. This smoothed histogram is then searched for the location of the first local maximum:

$$\widehat{\sigma}_{Ch} = \arg\max_{\sigma} \frac{1}{nh} \sum_{i=1}^{n} \kappa\left(\frac{\sigma - x_i}{h}\right). \tag{5.8}$$

5.2.2 New noise variance estimation method

In this subsection, a new noise variance estimation method will be described based on ML estimation.

Let $\{l_i\}$ with $i = 0, ..., B$ denote the set of boundaries of histogram bins. Furthermore, let n_i represent the number of observations (counts) within the bin $[l_{i-1}, l_i]$, which are multinomially distributed. Then, the joint PDF of the histogram data is given by [56]:

$$p(\{n_i\}|\sigma, \{l_i\}) = \frac{N_B!}{\prod_{i=1}^{B} n_i!} \prod_{i=1}^{B} p_i^{n_i}(\sigma) \quad , \tag{5.9}$$

with $N_B = \sum_{i=1}^{B} n_i$ the total number of observations within the partial histogram and p_i the probability that an observation assumes a value in the range $[l_{i-1}, l_i]$. For Rayleigh distributed observations, this probability is given by

$$p_i(\sigma) = \frac{\int_{l_{i-1}}^{l_i} \frac{m}{\sigma^2} \exp\left(-\frac{m^2}{2\sigma^2}\right) dm}{\sum_{i=1}^{B} \int_{l_{i-1}}^{l_i} \frac{m}{\sigma^2} \exp\left(-\frac{m^2}{2\sigma^2}\right) dm} \quad . \tag{5.10}$$

Since

$$\int_a^b \frac{m}{\sigma^2} \exp\left(-\frac{m^2}{2\sigma^2}\right) dm = e^{-\frac{a^2}{2\sigma^2}} - e^{-\frac{b^2}{2\sigma^2}} \quad , \tag{5.11}$$

it is easy to show that Eq. (5.10) simplifies to

$$p_i(\sigma) = \left(e^{-\frac{l_{i-1}^2}{2\sigma^2}} - e^{-\frac{l_i^2}{2\sigma^2}}\right) \left(e^{-\frac{l_0^2}{2\sigma^2}} - e^{-\frac{l_K^2}{2\sigma^2}}\right)^{-1} \quad . \tag{5.12}$$

When the set of observations $\{n_i\}$ is fixed and σ is regarded as a variable, the joint PDF given in Eq. (5.9) is called a likelihood function. The ML estimate is then found by maximizing this likelihood function L with respect to σ:

$$\widehat{\sigma}_{ML,B} = \arg\max_{\sigma} L(\sigma|\{n_i\}, \{l_i\}) \quad . \tag{5.13}$$

Equivalently, the ML estimate of σ is found by minimizing $-\ln L$ with respect to σ:

$$\widehat{\sigma}_{ML,B} = \arg\min_{\sigma} \left[N_B \ln\left(e^{-\frac{l_0^2}{2\sigma^2}} - e^{-\frac{l_K^2}{2\sigma^2}}\right) - \sum_{i=1}^{B} n_i \ln\left(e^{-\frac{l_{i-1}^2}{2\sigma^2}} - e^{-\frac{l_i^2}{2\sigma^2}}\right) \right]. \tag{5.14}$$

Eq. (5.14) is the ML estimator of the noise standard deviation σ from B bins.

This estimator Eq. (5.14) finds that specific value for the noise variance σ for which the probability Eq. (5.9) of observing the set of observations $\{n_i\}$ is maximal. That is, for any other value of σ the probability Eq. (5.9) of observing $\{n_i\}$ is lower.

Selection of the number of bins

Notice that basically the ML estimator Eq. (5.14) fits a (discretized) Rayleigh PDF to the partial (left side of the) MR image histogram. The criterion for the quality of the fit is given by the likelihood function. Since the number of bins B that are used is not given a-priori, this raises the question how the number of bins B is selected. Generally, a more precise estimate (i.e., a smaller variance) will be obtained if the number of bins taken into account increases, provided that the counts in those bins are indeed Rayleigh distributed background noise contributions. However, incorporating bins with counts that can not be attributed solely to noise but also to signal contributions will introduce a bias into the estimate of σ. Hence, as a selection criterion for B, a combined measure of the bias and variance of the estimator $\widehat{\sigma}_{ML,B}$ was chosen. This criterion is derived in the next paragraphs.

Variance A measure of the variance of $\widehat{\sigma}_{ML,B}$ was constructed from the Cramér-Rao lower bound (CRLB), which is a lower bound on the variance of any unbiased estimator $\widehat{\sigma}$ of σ [57]:

$$\mathbb{E}\left[(\sigma - \widehat{\sigma})^2\right] \geq \boldsymbol{I}^{-1}(\sigma) \quad , \tag{5.15}$$

with

$$\boldsymbol{I}(\sigma) = -\mathbb{E}\left[\frac{\partial^2}{\partial\sigma^2}\ln p(\{n_i\}|\sigma)\right] \tag{5.16}$$

the Fisher information, also known as the *expected* Fisher information. It is known that the ML estimator is consistent and asymptotically most precise (i.e., it attains the CRLB asymptotically). Therefore, a useful measure of the variance of $\widehat{\sigma}_{ML,B}$ is given by

$$\widehat{\mathrm{Var}}(\widehat{\sigma}_{ML,B}) = -\left(\frac{\partial^2}{\partial\sigma^2}\ln L(\sigma|\{n_i\})\Big|_{\sigma=\hat{\sigma}_{ML,B}}\right)^{-1} \quad . \tag{5.17}$$

The term right hand side is known as the inverse of the *observed* Fisher information. This estimate of the variance was observed to be reliable only when a sufficient number of bins was taken into account. In our implementation, this number was chosen such that at least the maximum of the histogram was included.

Bias A measure of the bias was found by quantifying the difference between the Rayleigh distribution fitted using the first B bins of the histogram and the actual bin counts in the histogram.

The histogram bin counts n_i are distributed with a multinomial distribution. Furthermore, the marginal distribution of the number of counts in each bin is a binomial distribution with parameters N_B and p_i. This means that the expected value of n_i is $p_i N_K$ and its variance is $p_i(1-p_i)N_B$. However, since in general N_B is large (and p_i is small), the binomial distribution can be approximated by a normal distribution with expectation value and variance both equal to $p_i N_K$. Under the null hypothesis (H_0) that the observations in all bins are Rayleigh distributed, p_i is given by Eq. (5.10). Next, consider the test statistic:

$$\lambda_K = \sum_{i=1}^{N} \frac{(f_{i,B} - n_i)^2}{f_{i,B}} \quad , \tag{5.18}$$

with N the number of bins in the histogram and

$$f_{i,B} = p_i(\widehat{\sigma}_{ML,B})N_B \quad . \tag{5.19}$$

It can be shown that, under H_0, λ_K is approximately χ^2_{N-2} distributed (i.e., chi-squared distributed with $N-2$ degrees of freedom). Obviously, H_0 is more likely to be rejected with increasing λ_K. Notice, that a large value of λ_K may indicate the presence of a bias in our estimate of σ. Therefore, λ_K will be used as a bias measure.

Most of the major contributions to λ_K can be expected to come from bins for which $i > B$, since these bins have not been taken into account in the estimation of σ. It is reasonable to assume that for these bins the counts due to the underlying, noiseless signal outnumber those due to the background noise only. Since contributions from the underlying signal can only increase the bin counts n_i, the actual bin counts will likely be significantly higher than the counts predicted by the fitted Rayleigh distribution. If we exclude the bins i with $i > B$ for which $n_i > f_{i,B}$ from Eq. (5.18), we obtain the modified test statistic:

$$\lambda^*_K = \sum_{i=1}^{B} \frac{(f_{i,B} - n_i)^2}{f_{i,B}} + \sum_{i=B+1}^{N} \frac{[\max(0, f_{i,B} - n_i)]^2}{f_{i,B}} \quad . \tag{5.20}$$

The first term of Eq. (5.20) is known as Pearson's test statistic [74], which is approximately (that is, asymptotically) χ^2 distributed with $B-2$ degrees of freedom under H_0. The second term of Eq. (5.20) is approximately χ^2_M distributed under H_0, with

$$M = \sum_{i=B+1}^{N} \epsilon(f_{i,B} - n_i) \quad . \tag{5.21}$$

Since both terms are independent, λ_K^* is approximately χ^2_{B-2+M} distributed under H_0. Hence, the statistic

$$\widehat{b} = \frac{\lambda_B^* - (B-2+M)}{\sqrt{B-2+M}} \tag{5.22}$$

has approximately a standard normal distribution under H_0. The statistic Eq. (5.22) will be used as a measure of the bias.

Selection criterion Finally, both measures of bias and variance given in Eq. (5.17) and Eq. (5.22), respectively, are combined into a single criterion that selects the optimal number of bins $\widehat{B}$:

$$\widehat{B} = \arg\min_K \left[\widehat{b} + \widehat{Var}(\widehat{\sigma}_{ML,B})\right]. \tag{5.23}$$

5.3 Experiments

Experiments were designed to compare the performance of the noise variance estimators discussed in subsection 5.2.1 to that of the newly proposed method presented in subsection 5.2.2. The experiments used simulated as well as experimental data. As a performance measure, the root-mean-squared-error (RMSE) was used.

Simulated noise-only images First, the performance of the estimators was compared using simulated, integer valued Rayleigh distributed data (corresponding to noise-only magnitude MR images), with different noise levels σ. The size of the image was 181×80.

Simulated three-modal image Next, an image was generated that would generate one background mode and two signal modes in the image histogram. In this way, overlap of the background mode with a signal mode could be studied. A three-modal image was obtained from an image with signal levels 0 (background), 100, and 200. Each level had an equal number of data points. Based on these levels, Rician distributed data were generated. Depending on σ, the modes overlapped which challenged estimation of the noise variance from the background mode. The size of the image was 181×240.

Simulated 2D MR image In a next experiment, a single slice of a noise free MR image was simulated using a web based MR simulator [58]. Thereby, the normal brain database was employed (Modality: T1 weighted; slice thickness: 3 mm; noise: 0%; Intensity non-uniformity (RF): 20%). Rician distributed data with varying σ were then generated from the noiseless image obtained from the simulator. The dimensions of the slice used were 181×217.

Simulated 3D MR image Next, a similar simulation experiment was set up as described above (i.e., using the web based MR simulator [58]), but now with a 3D MR image of size $181 \times 217 \times 60$.

Simulated 3D MR image with ghost Furthermore, the robustness of the noise variance estimators in the presence of a ghost artefact was tested. The ghost was generated by circularly shifting the original image in one direction over half the image size in that direction and scaling the intensities to 5% of the original intensities. This ghost was then added to the original image. Also for this simulation experiment, Rician distributed noise with different σ was added.

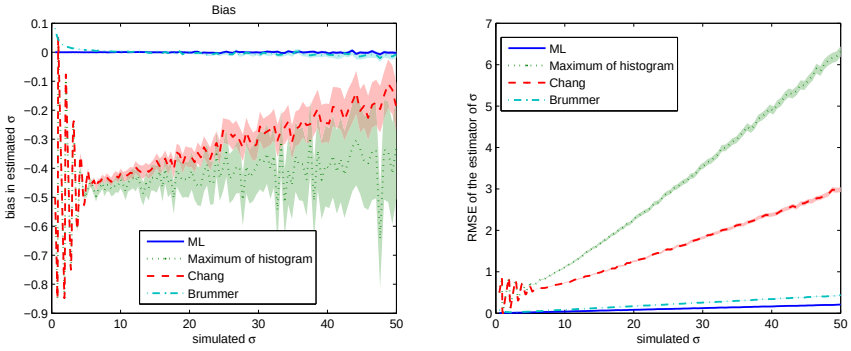
Experimental 3D MR images Finally, in order to test the different estimators on experimental data, a cherry tomato was scanned with a 7 Tesla (Bruker, DE) MR imaging system with self shielded gradients of 300 mT/m and an aperture of 10 cm.

To evaluate the standard deviation of the estimators experimentally, the estimators were applied to averaged images. Each averaged image was obtained by averaging over a number of images acquired under identical experimental conditions. Averaging was done in the complex k-space, so before reconstructing the magnitude image. The theoretical reduction of the noise standard deviation as a function of the number of images n over which the average was taken is known to be $1/\sqrt{n}$. Therefore, the estimated noise standard deviation, multiplied by $\sqrt{n}$ is expected to be constant as a function of n. In this experiment, it was tested whether the slope of the line obtained by linear regression differed significantly from zero.

5.4 Results and discussion

Simulated noise-only images In Fig. 5.2, the bias and RMSE of the different estimators are shown as a function of σ. At low noise levels, Chang's estimator and the Maximum estimator show an oscillatory behavior, which is caused by the discreteness of the histogram. Indeed, at low values of σ, the width of the Rayleigh distribution is smaller than the histogram bin width, which leads to an estimate of σ that is consistently located in the center of the bin, which in turn has a consistent negative or positive bias. Since for low σ, the smoothing parameter of Chang's estimator given by Eq. (5.6) is too small to compensate for this effect, the oscillatory behavior of this estimator is still apparent. For all values of σ, the Maximum estimator and Chang's estimator have significantly larger RMSE than Brummer's estimator and the ML based estimator.

Brummer's method and the ML based method account for the Rayleigh distribution, which leads to significantly improved RMSE values of the



(a) Bias (b) RMSE

Figure 5.2: *The bias (a) and RMSE (b) of the noise variance estimators as a function of σ for* ***simulated noise-only MR data****. For each value of σ, 5000 simulations were used.*

noise variance estimator. The proposed ML based noise variance estimator clearly performs best in terms of the RMSE because:

1. the ML based estimator correctly accounts for the discreteness of the data. This is especially important when σ is close to the histogram bin width. For all values of σ, only for the ML based estimator the bias could not be shown to be significantly different from zero (which can also be appreciated from Fig. 5.2a.
2. the multinomial distribution of the histogram bins is only taken into account by the ML based estimator. This results in a lower variance of the ML based estimator compared to that of Brummer's estimator for a given number of bins.
3. the number of bins to be used for estimation is adaptively determined. For noise only data, the ML based estimator takes generally all bins into account since they pass the Rayleigh distribution test (cfr. Eq. (5.20)) and thus has the lowest RMSE when the noise level is larger. In contrast, Brummer's method, the number of bins used for estimation is determined in a 'hard' way from an initial estimate of σ (cfr. Eq. (5.4).

The RMSE of the ML based estimator is approximately half of the RMSE of the second best, which is Brummer's estimator.

Simulated three-modal image In Fig. 5.4, the RMSE of the different estimators is plotted. As can be seen, the RMSE is low for most estimators when the signal level is below 1/3 of the first signal level and rises sharply

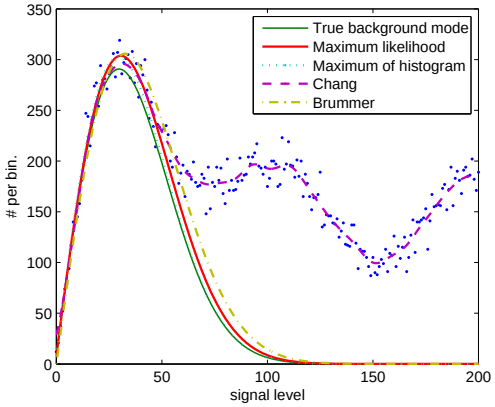


Figure 5.3: *Histogram of the* ***simulated three-modal-image*** *with standard deviation* $\sigma = 30$*, along with the true Rayleigh distribution as well as the Rayleigh distributions based on the estimated noise standard deviations and the low pass filtered histogram as specified by Chang's method.*

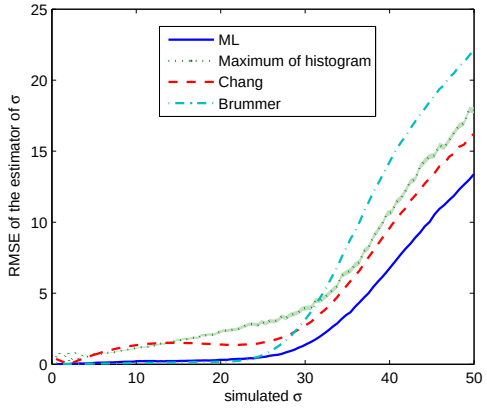
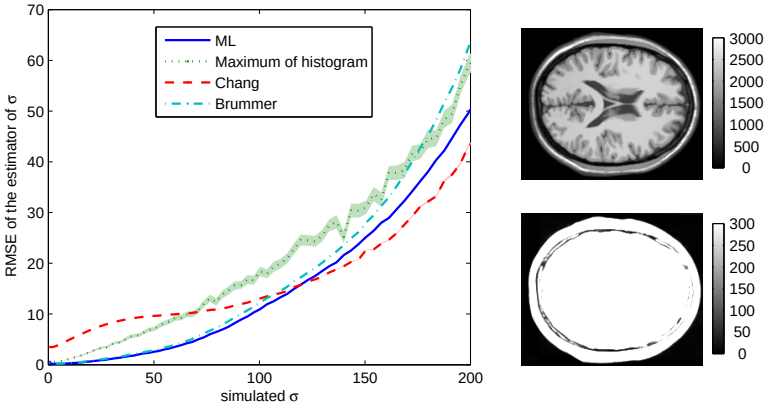


Figure 5.4: *The RMSE of the noise variance estimators as a function of* σ *for a* ***simulated three-modal MR image****. The simulated image contained three grey values: 0, 100, and 200. For each value of* σ*, 1000 simulations were used.*



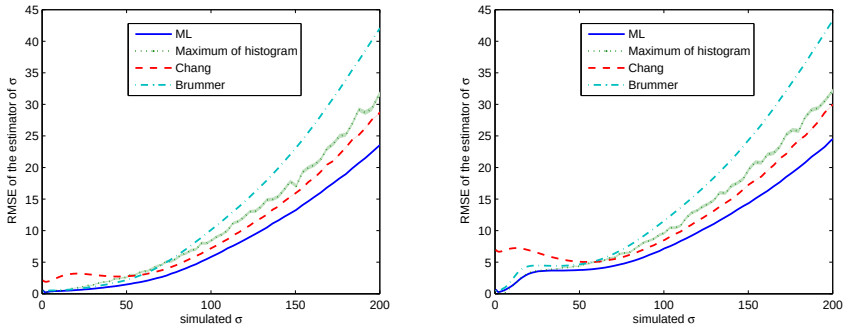
(a) (b)

Figure 5.5: *(a) The RMSE of the noise variance estimators as a function of σ for **simulated 2D MR data**. For each value of σ, 500 simulations were used. The noise free 2D slice used for this simulation is shown in (b).*

after that. For large σ (i.e., approximately $\sigma > 30$) the noise variance estimations yield less reliable results, because the background mode largely overlaps with the signal modes.

To illustrate the difficulty of estimating σ accurately, a representative realization of the histogram with a noise level of 30 is plotted in Fig. 5.3. Along with the histogram, the true, underlying Rayleigh distribution as well as the fitted Rayleigh distributions of the different estimators are shown. As can be observed, the fitted distribution using the proposed ML based estimation procedure, approximates the true distribution best. From Fig. 5.4, it is clear that for low σ (i.e., approximately $\sigma < 30$), both Brummer's method and the ML base method have significantly lower RMSE than the Maximum estimator and Chang's estimator, which is due to the fact that much more data from the histogram are taken into account, leading to a reduced variance of the noise variance estimator. For large σ (i.e., approximately $\sigma > 30$), the ML based estimator outperforms all other estimators with respect to the RMSE. This is because the ML based method tries to find the right balance between the variance and the bias of the σ estimator by optimizing the number of bins used for estimation.

Simulated 2D MR image The noise variance estimation results for simulated 2D MR image are shown in Fig. 5.5. Given that the mean value $\langle m \rangle$ of the noiseless image (in this case $\langle m \rangle$=210), the image SNR can be defined as $\langle m \rangle / \sigma$. For low SNR, the Chang's method performs best, prob-



(a) No Ghost (b) with Ghost

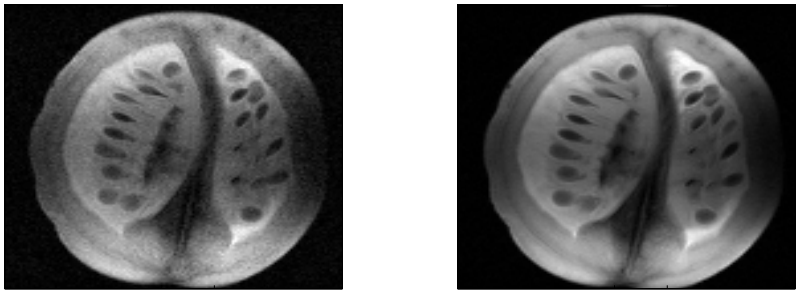
Figure 5.6: *The RMSE of the noise variance estimators as a function of σ for* ***simulated 3D-MR data****. The left image shows the results without ghost and the right image shows the results with a ghost added. For each value of σ, 500 simulations were used.*

ably caused by the smoothing of the histogram. For extremely low SNR, however, none of the methods are suitable for accurate noise variance determination because in this region the signal and noise contributions in the image histogram severely overlap. For moderate or high values of the SNR (i.e., SNR $>$ 2), the proposed ML based noise variance estimator performs best in terms of the RMSE.

Simulated 3D MR image The results of the simulated 3D data set are shown in Fig. 5.6a. For 3D data sets, the ratio of the number of background voxels to the number of non-background voxels is generally significantly larger compared to 2D data sets, which facilitates estimation of the noise variance from the histogram background mode.

In contrast to the noise-only data, Brummer's method scores worse for simulated 3D MR data than the Maximum and Chang's estimators. The main reason for this is that Brummer's estimator uses two times the initial noise σ estimate as the number of bins (cfr. Eq. (5.4)). When a lot of (background) data is present, as it is in a 3D image, the bias of this estimator becomes prominent. The ML based method, which searches for a compromise between precision and accuracy, uses fewer bins to obtain a lower RMSE value.

Simulated 3D MR image with ghost In Fig. 5.6b the results of the 3D image with ghost are presented. The change in the histogram of the noise free image which resulted from adding the ghost is mainly concentrated in the range 10 - 70. The ghost seems to slightly affect the noise variance



(a) No averaging (b) Average of 12 acquisitions

Figure 5.7: *MR image of a cherry tomato acquired with 1 and 12 images shown in (a) and (b), respectively.*

estimation for all noise variance estimation methods. However, also in this case, the proposed ML based estimator performs best in terms of the RMSE.

Experimental 3D MR images Finally, the noise variance was estimated from MR images of a cherry tomato. Fig. 5.7a and Fig. 5.7b show the MR reconstruction obtained by averaging over 1 and 12 acquired images, respectively. The resulting $\widehat{\sigma}$ as a function of n, for each estimator, is shown in Fig. 5.8. Chang's estimator did reveal a statistically significant trend, while the other estimators did not. Note that the variance of the Maximum estimator and Chang's estimator are larger than the variance of the ML based estimator and Brummer's estimator. This is because the latter estimators exploit a larger part of the Rayleigh distributed histogram background mode.

In general, we may conclude that the RMSE of the Maximum estimator performs worst of all described estimators in terms of the RMSE, mainly because the variance of this estimator is large. The RMSE of Chang's estimator is smaller than that of the Maximum estimator. However, in general, its RMSE is still significantly larger than that of Brummer's and the proposed ML based estimator. The large RMSE of the Maximum and Chang's estimators can partially be explained by the fact that they do not exploit the fact that the Rayleigh distribution characterizes background the histogram bins.

Brummer's method as well as the proposed ML based estimator do account for the Rayleigh distribution for the estimation of the noise variance. However, in general, the proposed ML estimator performs significantly better

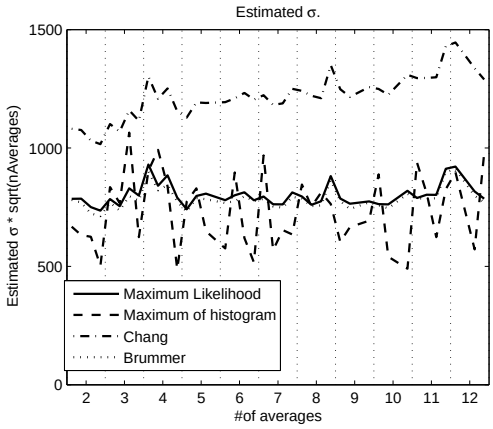


Figure 5.8: *Estimated σ of an experimental MR image of a cherry tomato, as a function of the number of averages n used during the acquisition.*

than Brummer's method, mainly because it selects the number of bins used to estimate the noise variance in an optimal way.

5.5 Conclusions

In this chapter, previously proposed noise variance estimation methods that employ the image histogram were reviewed and a new method was proposed based on Maximum Likelihood (ML) estimation. Simulation experiments showed that the ML based estimator outperforms the previously proposed estimators in terms of the root mean squared error.

Chapter 6

Likelihood based hypothesis tests for brain activation detection from FMRI data disturbed by colored noise

The main part of the work described in this chapter has been published as:

A. J. den Dekker, **D. H. J. Poot**, R. Bos and J. Sijbers

Likelihood based hypothesis tests for brain activation detection from FMRI data disturbed by colored noise: a simulation study

in: IEEE Transactions on Medical Imaging,
Volume 28, February 2009, Pages 287-296

Abstract

Functional magnetic resonance imaging (fMRI) data that are corrupted by temporally colored noise are generally preprocessed (i.e. prewhitened, or precolored) prior to functional activation detection. In this chapter, we propose likelihood based hypothesis tests that account for colored noise *directly* within the framework of functional activation detection.

Three likelihood based tests are proposed: the generalized likelihood ratio (GLR) test, the Wald test, and the Rao test. The fMRI time series is modeled as a linear regression model, where one regressor describes the task-related hemodynamic response, one regressor accounts for a constant baseline and one regressor describes potential drift. The temporal

correlation structure of the noise is modeled as an autoregressive (AR) model. The order of the AR model is determined from practical null data sets using Akaike's information criterion (with penalty factor 3) as order selection criterion. The tests proposed are based on exact expressions for the likelihood function of the data.

Using Monte Carlo simulation experiments, the performance of the proposed tests is evaluated in terms of detection rate and false alarm rate properties and compared to the current general linear model (GLM) test, which estimates the coloring of the noise in a separate step. Results show that theoretical asymptotic distributions of the GLM, GLR, and Wald test statistics cannot be reliably used for computing thresholds for activation detection from finite length time series. Furthermore, it is shown that, for a fixed false alarm rate, the detection rate of the proposed GLR test statistic is slightly, but (statistically) significantly improved compared to that of the common GLM based tests. Finally, simulations results reveal that all tests considered show seriously inferior performance if the order of the AR model is not chosen sufficiently high to give an adequate description of the correlation structure of the noise, whereas the effects of (slightly) overmodeling are observed to be less harmful.

keywords fMRI, Statistical Parametric Maps, Generalized Likelihood Ratio test, time series analysis, Wald test, Rao test

6.1 Introduction

Functional magnetic resonance imaging (fMRI) is a noninvasive technique used to detect brain activity. By utilizing the fact that the magnetic resonance signal intensity is correlated with the cerebral blood flow, which in turn is correlated with neural activity [4], fMRI can localize brain regions that show significant neural activity upon stimulus presentation, where the stimulus is a task designed to activate specific brain regions related to the studied cognitive process. fMRI data sets typically consist of time series associated with the voxels of the brain. For each voxel, the significance of the response to the stimulus is assessed by statistically analyzing the associated fMRI time series. In this way, brain activation maps, or statistical parametric maps (SPMs), reflecting brain activity can be constructed.

Nowadays, fMRI time series are commonly modeled by a general linear model (GLM) disturbed by Gaussian distributed noise [59, 60]. Such a model is capable of including potential time trends by adopting extra linear terms. Furthermore, the GLM contains one or more activation related parameters of interest. SPMs are obtained by testing the significance of the activation related GLM parameter(s) using standard statistical tools such as the two-sided t-test (in the one parameter case) or the F-test (in the case of more than

one parameter). This method is also used in the 'Estimate' and 'Inference' steps of the well known SPM software package [61].

The fMRI recordings are contaminated by noise from sources such as the MRI scanner, residual motion, physiological processes (breathing, blood circulation), and non task related 'spontaneous' brain activations [3]. It is reasonable to assume that this noise is colored (i.e. correlated) in the time direction. Unlike white noise, colored noise does not have a uniform (i.e. flat) power spectral density function. Since the underlying correlation structure is unknown, current methods deal with temporally correlated noise by prewhitening the data based on an *estimated* correlation matrix of the noise [60]. This correlation matrix is usually estimated by fitting an autoregressive (AR) time series model to the residuals obtained after fitting the general linear model to the fMRI time series in least squares sense [62]. This introduces a, usually small, bias in the correlation estimates [63]. Since an estimate of the correlation matrix instead of the unknown, true correlation matrix of the noise is used for prewhitening, the assumption that the test statistic has a Student's t or F distribution (upon which inference on the significance of the response is based) is only approximately valid.

In this chapter, an alternative approach is proposed. This approach is also based on a GLM with correlated noise modeled as an AR process, but unlike the common GLM approach, it does not require a prewhitening step. Instead, statistical inference is based on the exact likelihood function that describes the statistics of the data including the temporal correlation structure of the noise. No approximations are made. The order of the AR process is determined from practical null data sets, acquired in the absence of activity. Three likelihood based statistical binary hypothesis tests are proposed: the generalized likelihood ratio test (GLRT), the Wald test, and the Rao test. In each case, the null hypothesis H_0 that no activation is present, is tested against the alternative hypothesis H_1 (activation is present). In the context of fMRI, the use of the GLRT has previously been proposed by Nan and Nowak [64]. However, they consider complex valued fMRI data contaminated with white noise while in the present work, we consider magnitude fMRI data and colored noise.

For the computation of the test statistics proposed, the maximum likelihood (ML) estimates of the unknown parameters under H_0 (Rao), H_1 (Wald), or both H_0 and H_1 (GLRT) are needed. They are obtained by maximizing the likelihood function with respect to all unknown parameters (including the parameters of the AR model) simultaneously. In this chapter, the performance of the proposed tests are evaluated in terms of detection rate and false alarm rate.

It is known that the tests proposed have favorable asymptotic statistical properties [65]. The asymptotic statistical distributions of the test statistics under H_0 do not depend on any unknown parameters. Therefore, independent of the noise power, tests can be constructed that have a (specified) constant

false alarm rate. Such a test is referred to as a constant false alarm rate (CFAR) test [65]. Whether these asymptotic properties also apply to a finite number of observations is investigated by means of simulation experiments. The performance of the proposed tests is also compared to that of the widely used t-test (which is based on the GLM approach).

The remaining part of this chapter is organized as follows. Section 6.2.1 describes a general, statistical model of fMRI time series. Section 6.2.2 discusses the general linear model (GLM) approach, assuming correlated noise described by an autoregressive (AR) process. In Section 6.2.3, the joint probability density function (PDF) of the data is derived. In Section 6.2.4, some optimizations are introduced to efficiently compute the ML estimate. The Sections 6.2.5, 6.2.6, 6.2.7, 6.2.8 and 6.2.9 describe the different test statistics. In Section 6.3, experimental results are described. Section 6.3.1 describes how to determine the order of the AR process from null data sets. In the Section 6.3.2 and Section 6.3.3, the tests are applied to simulated and experimental data.

Finally, conclusions are drawn in Section 6.4.

6.2 Method

6.2.1 The statistical model of the fMRI time series

An fMRI time series $\boldsymbol{y} = [y_1, \ ..., \ y_n]^T$ (the superscript T denotes matrix transposition) of equidistant observations can in general be modelled as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{v} \quad , \tag{6.1}$$

in which $\boldsymbol{X}$ is an $n \times m$ design matrix [3,59]. It consists of m columns that model signals of interest and nuisance signals such as potential drift. Furthermore, $\boldsymbol{\theta}$ is an $m \times 1$ vector of unknown regression parameters and $\boldsymbol{v}$ is an $n \times 1$ vector that represents stochastic noise contributions. The noise $\boldsymbol{v}$ is modelled as a stationary stochastic AR process of order r (i.e. an AR(r) process) [66]:

$$v_t = e_t - \alpha_1 v_{t-1} - \alpha_2 v_{t-2} - \cdots - \alpha_r v_{t-r} \quad , \tag{6.2}$$

with $\boldsymbol{\alpha} = [\alpha_1, \ \ldots, \ \alpha_r]^T$ the vector of AR parameters, t the time index and e_t independent, zero mean Gaussian distributed white noise with variance σ_e^2. Let $\sigma_e^2 \boldsymbol{V}$ be the $n \times n$ covariance matrix of the AR process Eq. (6.2), that is,

$$\sigma_e^2 \boldsymbol{V} = \mathbb{E}\left[\boldsymbol{v}\boldsymbol{v}^T\right] \quad , \tag{6.3}$$

with $\boldsymbol{v} = (v_1, \ldots, v_n)^T$ and $\mathbb{E}[.]$ the expectation operator. For observations of stationary stochastic processes, the covariance matrix has a Toeplitz structure.

Therefore, the covariance matrix of the AR(r) process v_t may be written as

$$\sigma_e^2 \boldsymbol{V} = \sigma_v^2 \begin{bmatrix} \rho(0) & \rho(1) & \dots & \rho(n-1) \\ \rho(1) & \rho(0) & \dots & \rho(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(n-1) & \rho(n-2) & \dots & \rho(0) \end{bmatrix} \quad , \tag{6.4}$$

where $\rho(k) = \mathbb{E}[v_t v_{t+k}]/\sigma_v^2$ and σ_v^2 is the variance of v_t. Notice that it follows from this definition that $\rho(0) = 1$. The elements of the matrix $\boldsymbol{V}$ can be expressed in terms of the AR parameters through the $(r+1)$ Yule Walker relations [**?**]:

$$\rho(k) + \alpha_1 \rho(|k-1|) + \cdots + \alpha_r \rho(|k-r|) = 0 \quad \forall k \in \{0 \dots r\}. \tag{6.5}$$

Furthermore, it can be shown that [66]

$$\sigma_v^2 = \frac{\sigma_e^2}{1 - \sum_{k=1}^{r} \alpha_k \rho(k)} \quad . \tag{6.6}$$

Several authors have performed analyzes that indicate that AR models give an accurate description of the actual temporal autocorrelation structure of the noise that contaminates fMRI data [**?**, 62]. The validity of the model will be assessed using experimental data in Section 6.3.1.

In order to derive the different test statistics in the Sections 6.2.6, 6.2.7, 6.2.8 and 6.2.9, first the generalized least squares (GLS) estimator, the joint probability density function of the data and the ML estimator are derived in the Sections 6.2.2, 6.2.3 and 6.2.4, respectively.

6.2.2 The common GLM approach

The widely used GLM approach, for example by SPM [61], consists of two steps. First, an estimate of the parameter vector $\boldsymbol{\theta}$ is obtained by least squares fitting of the model described by the right hand side of Eq. (6.1) to the data $\boldsymbol{y}$. This so-called ordinary least squares (OLS) estimator can be expressed in closed form by

$$\widehat{\boldsymbol{\theta}}_{OLS} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y} \quad . \tag{6.7}$$

Although not fully efficient, this estimator is unbiased [57]. Therefore, the residuals $\boldsymbol{\varepsilon}_{OLS} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}_{OLS}$ have zero expectation values and a correlation structure that is approximately equal to that of the noise $\boldsymbol{v}$. Assuming that the noise is generated by an AR(r) model, the parameters of this model and hence the matrix $\boldsymbol{V}$ can be estimated from the residuals [60]. In the simulation experiments, described in Section 6.3 of this chapter, the `sig2ar` function of the ARMASA Matlab toolbox [**?**], was used for this estimation. The estimated covariance matrix will be denoted as $\widehat{\boldsymbol{V}}$.

Second, $\widehat{\boldsymbol{V}}^{-1}$ is used as weighting matrix in a generalized least squares (GLS) estimator of $\boldsymbol{\theta}$, which results in

$$\widehat{\boldsymbol{\theta}}_{GLS} = (\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{y} = \widehat{\boldsymbol{W}}\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{y} \quad , \tag{6.8}$$

where the $m \times m$ matrix $\widehat{\boldsymbol{W}} = (\boldsymbol{X}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}$ is an estimator of the covariance matrix of $\widehat{\boldsymbol{\theta}}_{GLS}$. Notice that estimator Eq. (6.8) is equivalent to prewhitening the data and model with $\widehat{\boldsymbol{V}}^{-1/2}$. That is, with $\tilde{\boldsymbol{X}} = \widehat{\boldsymbol{V}}^{-1/2}\boldsymbol{X}$ and $\tilde{\boldsymbol{y}} = \widehat{\boldsymbol{V}}^{-1/2}\boldsymbol{y}$ the GLS estimator can be written as

$$\widehat{\boldsymbol{\theta}}_{GLS} = (\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{y}} \tag{6.9}$$

In principle, the procedure can be iterated by repeating both steps described above, that is, by re-estimating the covariance matrix $\boldsymbol{V}$ from the residuals

$$\boldsymbol{\varepsilon}_{GLS} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}_{GLS} \tag{6.10}$$

and substituting the result in Eq. (6.8). However, this procedure was not implemented in the simulation experiments described in Section 6.3, since it was observed that iterating didn't change the results significantly. Notice that if $\boldsymbol{V}$ is known, an unbiased estimator of σ_e^2 is given by

$$\widetilde{\sigma_e^2} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y})^T\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y})/(n-m) \quad , \tag{6.11}$$

where $\boldsymbol{W} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ and $(n-m)\widetilde{\sigma_e^2}/\sigma_e^2$ is χ^2 distributed with $n-m$ degrees of freedom. If we substitute the estimator $\widehat{\boldsymbol{V}}$ for $\boldsymbol{V}$ in Eq. (6.11), we yield the estimator

$$\widehat{\sigma_e^2} = \boldsymbol{\varepsilon}_{GLS}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{\varepsilon}_{GLS}/(n-m) \tag{6.12}$$

of which the statistics are not known exactly. However, as we will see later, the validity of the assumption that the test statistic Eq. (6.32) associated with the widely used F-test (described in Section 6.2.6) has indeed an F-distribution is subject to the validity of the assumption that $\widehat{\sigma_e^2}$ has the same distribution as estimator $\widetilde{\sigma_e^2}$. Obviously, this assumption is questionable.

Note that the GLM method described above can be implemented for any AR model order on a voxel by voxel basis. This differs from its implementation in the SPM software package [61], where only a single, iteratively estimated, global AR(1) model for all brain voxels is used.

6.2.3 The joint probability density function of the data

In order to derive the ML estimator of $\boldsymbol{\theta}, \boldsymbol{\alpha}$ and σ_e^2, the joint probability density function (PDF) of the fMRI data is needed. This joint PDF $p(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2)$ can be factorized as:

$$p(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2) = p(\boldsymbol{y}_{r+1:n}|\boldsymbol{y}_{1:r}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2)\, p(\boldsymbol{y}_{1:r}; \boldsymbol{\theta}, \boldsymbol{\alpha}, \sigma_e^2), \tag{6.13}$$

with $\boldsymbol{y}_{1:r} = [y_1, \ldots, y_r]^T$ and $\boldsymbol{y}_{r+1:n} = [y_{r+1}, \ldots, y_n]^T$. With Eq. (6.1) and Eq. (6.2) it can be shown that

$$e_t = y_t - \boldsymbol{x}_t\boldsymbol{\theta} + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}\boldsymbol{\theta}) + \alpha_r(y_{t-r} - \boldsymbol{x}_{t-r}\boldsymbol{\theta}), \tag{6.14}$$

where $\boldsymbol{x}_t$ denotes the t-th row of the design matrix $\boldsymbol{X}$. Therefore, the conditional PDF of the observations $\boldsymbol{y}_{r+1:n}$, given that the first r observations $\boldsymbol{y}_{1:r}$ remain fixed at their observed values, may be written as [66, p. 347]

$$p(\boldsymbol{y}_{r+1:n}|\boldsymbol{y}_{1:r};\boldsymbol{\theta},\boldsymbol{\alpha},\sigma_e^2) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{(n-r)/2} \times \tag{6.15}$$

$$\exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=r+1}^{n}\{y_t - \boldsymbol{x}_t\boldsymbol{\theta} + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}\boldsymbol{\theta}) + \ldots + \alpha_r(y_{t-r} - \boldsymbol{x}_{t-r}\boldsymbol{\theta})\}^2\right).$$

The joint PDF of the data $\boldsymbol{y}_{1:r}$ may be written as [66, p. 350]

$$\begin{aligned} p(\boldsymbol{y}_{1:r};\boldsymbol{\theta},\boldsymbol{\alpha},\sigma_e^2) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{r/2} |\boldsymbol{V}_r|^{-1/2} \times \\ \exp\left(-\frac{1}{2\sigma_e^2}(\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta})^T \boldsymbol{V}_r^{-1}(\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta})\right), \end{aligned} \tag{6.16}$$

where $\boldsymbol{X}_{1:r}$ denotes the $r \times m$ matrix consisting of the first r rows of the design matrix $\boldsymbol{X}$. Furthermore, $\boldsymbol{V}_r$ denotes the $r \times r$ covariance matrix of $\boldsymbol{v}_{1:r} = (v_1, \ldots, v_r)^T$ divided by σ_e^2 and $|\boldsymbol{V}_r|$ denotes the determinant of $\boldsymbol{V}_r$. By multiplying the conditional PDF in Eq. (6.15) by Eq. (6.16), the exact joint PDF of the data $\boldsymbol{y}$ may be written as [66]

$$p(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha},\sigma_e^2) = \left(\frac{1}{2\pi\sigma_e^2}\right)^{n/2} |\boldsymbol{V}_r|^{-1/2} \exp\left(-\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})/2\sigma_e^2\right), \tag{6.17}$$

where

$$\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}) = (\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta})^T \boldsymbol{V}_r^{-1}(\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta}) + Q(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\alpha}) \tag{6.18}$$

and

$$Q(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}) = \sum_{t=r+1}^{n}\{y_t - \boldsymbol{x}_t\boldsymbol{\theta} + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}\boldsymbol{\theta}) + \ldots + \alpha_r(y_{t-r} - \boldsymbol{x}_{t-r}\boldsymbol{\theta})\}^2 \tag{6.19}$$

are defined for convenience.

6.2.4 Maximum likelihood estimator

When the data $\boldsymbol{y}$ are given, the PDF given in Eq. (6.17) is a function of the parameters $\boldsymbol{\alpha}, \boldsymbol{\theta}$ and σ_e^2 only and it is called the likelihood function. In order

to compute the likelihood based tests, the ML estimate of the unknown parameters has to be found, both under the null hypothesis H_0 and the alternative hypothesis H_1. For that purpose, the likelihood function has to be maximized with respect to the unknown parameters $(\boldsymbol{\alpha}, \boldsymbol{\theta}, \sigma_e^2)$. Note that maximization of the likelihood function is equivalent to maximization of the (natural) logarithm of the likelihood function because the logarithmic function is monotomic. It follows from Eq. (6.17) that the natural logarithm of the likelihood function, which is called the log-likelihood function, is given by

$$\ln\left(p(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha},\sigma_e^2)\right) = -\frac{n}{2}\ln\left(2\pi\sigma_e^2\right) - \frac{1}{2}\ln\left(|\boldsymbol{V}_r|\right) - \frac{1}{2\sigma_e^2}\widetilde{Q}\left(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}\right). \tag{6.20}$$

The noise variance σ_e^2 can be eliminated from the optimization problem since the value of σ_e^2 that maximizes the likelihood function $p(\boldsymbol{y};\boldsymbol{\alpha},\boldsymbol{\theta},\sigma_e^2)$ can easily be solved from

$$\frac{\partial \ln p(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha},\sigma_e^2)}{\partial\sigma_e^2} = -\frac{n}{2\sigma_e^2} - \frac{1}{2\sigma_e^4}\widetilde{Q}\left(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}\right) = 0 \quad , \tag{6.21}$$

and is equal to

$$\hat{\sigma}_e^2 = \frac{1}{n}\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}). \tag{6.22}$$

Substituting Eq. (6.22) in Eq. (6.17) yields the so-called concentrated likelihood function:

$$p(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}) = \left(\frac{n}{2\pi\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})}\right)^{n/2} |\boldsymbol{V}_r|^{-1/2}\exp\left(-\frac{n}{2}\right) \quad . \tag{6.23}$$

Notice that $\boldsymbol{V}_r$ depends on the parameters $\boldsymbol{\alpha}$. The ML estimates $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\theta}})$ of the parameters $(\boldsymbol{\alpha}, \boldsymbol{\theta})$ can now be found by maximizing Eq. (6.23) with respect to $(\boldsymbol{\alpha}, \boldsymbol{\theta})$, both with and without the H_0 constraints. The maximization of the likelihood function is a nonlinear optimization problem that can be solved numerically. Substituting $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\theta}})$ for $(\boldsymbol{\alpha}, \boldsymbol{\theta})$ in Eq. (6.22) yields the ML estimate of σ_e^2.

For computational reasons, the logarithm of the concentrated likelihood function

$$\ln\left(p(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})\right) = -\frac{n}{2}\ln\left(\frac{2\pi}{n}\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})\right) - \frac{1}{2}\ln\left(|\boldsymbol{V}_r|\right) - \frac{n}{2}\ , \tag{6.24}$$

is maximized. Since numerical optimization is much more efficient when the gradient (and hessian) is available, one can also compute the first (and second) derivative of the concentrated log-likelihood function with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$. These derivatives, in which non trivial simplifications have been applied, are explained in Appendix 6.A.

In the simulation experiments described in Section 6.3 of this chapter, the ML estimator was implemented in MATLAB, using a built-in unconstrained optimization routine which uses a subspace trust region method and is based on the interior-reflective Newton method [**?**].

Extension to multiple independent traces

With the above method, a noise model is computed for each voxel trace. However, analyzing measured brain data indicated that the correlation structures of the noise of neighboring voxels are similar. Since the power of inferences usually increases with increasing number of data points available for estimation, it would be beneficial to estimate a single noise model from multiple neighboring traces. We observed that in our non-motion-corrected fMRI data sets the spatial correlation of the noise was quite low. Thus, the correlation structure is similar between neighboring voxels, while the actual noise realizations are (almost) uncorrelated. This would justify simultaneous modeling of multiple neighboring voxel traces with a single AR process. As an extension of Eq. (6.23), in this section, the time series of neighboring voxels are treated as independent and are modeled by one single AR process describing the coloring of the noise of the set of neighboring voxels. Let $\boldsymbol{Y}$, $\boldsymbol{Y}_i = \boldsymbol{y}$ with $i \in \{1 \dots M\}$, be the set of the time series of the M voxels in the region of which the noise can be described by one noise model. Furthermore, let $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}_i = \boldsymbol{\theta}$, be the set of regression parameter vectors from these M voxels: Then, the joint PDF is given by

$$\begin{aligned}\ln p(\boldsymbol{Y}|\boldsymbol{\Theta}, \boldsymbol{\alpha}, \sigma_e^2) &= \sum_{i=1}^{M} -\frac{n}{2}\ln\left(2\pi\sigma_e^2\right) - \frac{1}{2}\ln\left(|\boldsymbol{V}_r|\right) - \frac{1}{2\sigma_e^2}\widetilde{Q}\left(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i, \boldsymbol{\alpha}\right) \\ &= -\frac{Mn}{2}\ln\left(2\pi\sigma_e^2\right) - \frac{M}{2}\ln\left(|\boldsymbol{V}_r|\right) - \frac{1}{2\sigma_e^2}\sum_{i=1}^{M}\widetilde{Q}\left(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i, \boldsymbol{\alpha}\right) \quad . \end{aligned} \tag{6.25}$$

The σ_e^2 that maximizes this likelihood is given by

$$\hat{\sigma}_e^2 = \frac{1}{nM}\sum_{i=1}^{M}\widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i, \boldsymbol{\alpha}) \quad , \tag{6.26}$$

with which again a concentrated likelihood can be computed

$$\ln\left(p(\boldsymbol{\Theta}, \boldsymbol{\alpha}|\boldsymbol{Y})\right) = -\frac{nM}{2}\left[1 + \ln\left(\frac{2\pi}{nM}\sum_{i=1}^{M}\widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i, \boldsymbol{\alpha})\right)\right] - \frac{M}{2}\ln\left(|\boldsymbol{V}_r|\right) \quad . \tag{6.27}$$

As both the noise level and coloring of the noise is estimated from M voxel time series, it is expected that they have a substantially improved precision and the strength of the activation tests might be improved.

6.2.5 Statistical inference

Brain activation can now be detected on a voxel by voxel basis by testing the significance of the task-related parameter(s). To determine whether a voxel is active or not, one can distinguish one-sided and two-sided tests. One-sided tests should be applied when the sign of the activation parameter(s) is known a-priori. Since this is usually not justified in fMRI experiments [67], we will restrict our analysis to two-sided tests. However, the methods presented can be easily extended to one-sided tests. Since all considered tests can easily be stated for the multi parameter case, this section is not restricted to single parameter testing, but to a more general linear hypothesis test. Suppose that we wish to test if $\boldsymbol{\tau}$ satisfies the linear equations $\boldsymbol{C\tau} = \boldsymbol{c}$, where $\boldsymbol{C}$ is a known full rank $j \times (m+r+1)$ matrix and $\boldsymbol{c}$ is a known $j \times 1$ vector. Then a two-sided hypothesis test can be specified by

$$H_0 : \boldsymbol{C\tau} = \boldsymbol{c} \tag{6.28}$$

$$H_1 : \boldsymbol{C\tau} \neq \boldsymbol{c}. \tag{6.29}$$

The hypothesis test decides H_0 when $\boldsymbol{C\tau}$ is not statistically significantly different from $\boldsymbol{c}$ and H_1 otherwise. For testing the presence of activation, usually $j = 1$, $\boldsymbol{c} = \boldsymbol{0}$, and $\boldsymbol{C}$ reduces to a row vector in which only the element corresponding to the activation parameter (e.g. θ_1) is nonzero. For some tests, the ML estimator of the parameters with and/or without the constraints imposed by Eq. (6.28) are needed. When we substitute the values of the acquired data $\boldsymbol{y}$ in the expression for the joint PDF of the data, given by Eq. (6.17), the resulting function is a function of the unknown parameters $\boldsymbol{\tau}$ only. The ML estimates under H_0 and H_1 are then given by, respectively,

$$\widehat{\boldsymbol{\tau}_0} = \arg\max_{\boldsymbol{\tau}} p\left(\boldsymbol{y}|\boldsymbol{\tau}\right), \text{ subject to } \boldsymbol{C\tau} = \boldsymbol{c} \tag{6.30}$$

$$\widehat{\boldsymbol{\tau}_1} = \arg\max_{\boldsymbol{\tau}} p\left(\boldsymbol{y}|\boldsymbol{\tau}\right). \tag{6.31}$$

In the next section, the GLM based F-test is reviewed. Subsequently, three likelihood based tests are described: the GLRT, the Rao test, and the Wald test. All these tests are based on the joint PDF of the data, described in Section 6.2.3. Furthermore, the Rao test and the Wald test are based on the Fisher information matrix, derived in the appendix.

6.2.6 F-test

For the GLM based F test, where only the linear regression parameters $\boldsymbol{\theta}$ can be tested, $\boldsymbol{\tau} = \boldsymbol{\theta}$ and $\boldsymbol{C}$ is a $j \times m$ matrix. The test statistic of the F-test is

then given by

$$T_F = \frac{\left(\boldsymbol{C}\widehat{\boldsymbol{\theta}}_{\mathrm{GLS}} - \boldsymbol{c}\right)^T \left(\boldsymbol{C}\widehat{\boldsymbol{W}}\boldsymbol{C}^T\right)^{-1} \left(\boldsymbol{C}\widehat{\boldsymbol{\theta}}_{\mathrm{GLS}} - \boldsymbol{c}\right)}{\widehat{\sigma_e^2}}, \tag{6.32}$$

where $\widehat{\boldsymbol{\theta}}_{GLS}$ denotes the GLS estimator and $\widehat{\sigma_e^2}$ is given by Eq. (6.12). Under H_0, the test statistic T_F has approximately an F distribution with j and $n-m$ degrees of freedom. If $\boldsymbol{V}$ would be known, T_F would be exactly F distributed with the specified degrees of freedom. The F-test, decides H_1 if $T_F > \gamma$, with γ some user specified threshold. This threshold is usually computed using the F distribution and balancing the false alarm rate (probability of deciding H_1 when H_0 is true) against the detection rate (probability of deciding H_1 when H_1 is true).

6.2.7 The generalized likelihood ratio test (GLRT)

The generalized likelihood ratio (GLR) is given by [65]:

$$\lambda = \frac{p\left(\boldsymbol{y}|\widehat{\boldsymbol{\tau}}_0\right)}{p\left(\boldsymbol{y}|\widehat{\boldsymbol{\tau}}_1\right)} \quad . \tag{6.33}$$

The GLRT principle now states that H_0 is to be rejected if and only if $\lambda \leq \lambda_0$, where λ_0 is some user specified threshold. It can be shown that, asymptotically (i.e. for $n \to \infty$), the modified GLR test statistic

$$T_{LR} = -2\ln\lambda \tag{6.34}$$

possesses a χ^2_j distribution with j degrees of freedom when H_0 is true.

Notice that, if $\boldsymbol{X}$ is not of full rank, the optimization implicit in Eq. (6.33) might be numerically difficult. In that case, the use of Bayes factors may be considered [**?**].

6.2.8 The Wald test

The Wald test statistic is given by [65]

$$T_{W2} = \left(\boldsymbol{C}\widehat{\boldsymbol{\tau}}_1 - \boldsymbol{c}\right)^T \left(\boldsymbol{C}\boldsymbol{F}^{-1}(\widehat{\boldsymbol{\tau}}_1)\boldsymbol{C}^T\right)^{-1} \left(\boldsymbol{C}\widehat{\boldsymbol{\tau}}_1 - \boldsymbol{c}\right), \tag{6.35}$$

where $\boldsymbol{F}^{-1}(\widehat{\boldsymbol{\tau}}_1)$ is the inverse of the Fisher information matrix (see Appendix 6.B), evaluated at $\widehat{\boldsymbol{\tau}}_1$. The (two-sided) Wald test decides H_1 if $T_{W2} > \gamma$, where γ is some user specified threshold. Asymptotically, the test statistic T_{W2} has a χ^2_j distribution, that is, a χ^2 distribution with j degrees of freedom under H_0.

6.2.9 The Rao test

The Rao test statistic is given by [65]

$$T_{R2} = \frac{\partial \ln p(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\bigg|^{T}_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}_0} \boldsymbol{R}\boldsymbol{C}\boldsymbol{F}^{-1}(\widehat{\boldsymbol{\tau}}_0)\boldsymbol{C}^T\boldsymbol{R}^T \frac{\partial \ln p(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}\bigg|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}_0} \tag{6.36}$$

where $\boldsymbol{F}^{-1}(\widehat{\boldsymbol{\tau}}_0)$ is the inverse of the Fisher information matrix, evaluated at $\boldsymbol{\tau} = \widehat{\boldsymbol{\tau}}_0$ and the $(m + r + 1) \times j$ matrix $\boldsymbol{R}$ is the pseudoinverse of $\boldsymbol{C}$. The Rao test decides H_1 if $T_{R2} > \gamma$, where γ is some user specified threshold. Asymptotically, the test statistic T_{R2} has a χ^2_j distribution under H_0.

6.2.10 Discussion

Knowledge of the PDF of the test statistic under H_0 allows one to compose tests with a desired *false alarm rate*. The false alarm rate is the probability that the test will decide H_1 when H_0 is true. The *detection rate* is the probability that the test will decide H_1 when H_1 is true. Throughout this chapter, we will denote the false alarm rate and the detection rate by P_f and P_d, respectively. Furthermore, a test has the so-called constant false-alarm rate (CFAR) property if the threshold required to maintain a constant P_f can be found independently of the signal-to-noise ratio (SNR) [65], which is usually unknown beforehand. Since the asymptotic PDFs of the likelihood based test statistics discussed in this section are known and parameter and SNR invariant, the tests will all have the CFAR property at least asymptotically. Whether or not the tests have the CFAR property for a finite number of observations can be found out by means of simulations. For more details on likelihood based tests, see [65].

6.3 Experiments and results

Experimental fMRI data sets were obtained from a healthy human volunteer, male, age 32 years. An informed consent was signed by the participant. All human experiments were performed on a 1.5T scanner with high-performance 40 mT/m gradients (Siemens Sonata, Erlangen, Germany). Gradient-recalled multi-shot EPI sequences (TE 50 ms, TR 3000 ms) were used with 30 slices covering the whole brain. The voxel dimensions were 3mm×3mm×3mm. Head movement was restricted by foam-padded cushions and the subject wore ear-plugs and noise reducing head-phones throughout the entire experiment.

6.3.1 Estimation of the order of the AR and trend models

Experimental fMRI resting (null) data were used to determine relevant orders of the AR noise process and trend model.

The trend model we used was a polynomial of order m (to be selected):

$$\text{trend} = \Sigma_{k=0}^{m} b_k t^k. \tag{6.37}$$

The noise was modeled by an AR process as in Eq. (6.2). AR models of orders 0 to 8 and trend models of orders 0 to 4 were evaluated for a random selection of 10 000 brain voxel traces with 90 time points from an fMRI null data set. The polynomial order and AR order of each voxel was selected using Akaike's information criterion (AIC) [**?**], where a penalty factor of 3 instead of 2 was chosen [**?**],

$$\text{AIC}_3 = 2 \ln p(\boldsymbol{y}|\widehat{\boldsymbol{\tau}}_1) + 3(m + r). \tag{6.38}$$

A histogram of the selected orders is plotted in Fig. 6.1a. For most traces, the selected order of the polynomial was 0 or 1 (linear trend). Also, for most traces the AR order selected was between 0 and 4. Due to the statistics involved in the order selection, it is unlikely that for all traces the selected order of the model equals the model of the underlying process. In order to get an impression of the orders selected for a given model, a simulation of AR(4) noise with a linear trend was set up. The simulation also had 90 time points per voxel trace. The parameters of the AR noise generating process and the trend used are given by

$$\begin{aligned} \mathbf{AR} &= [1 \quad -0.177 \quad -0.164 \quad -0.115 \quad -0.130] \\ \text{trend} &= 0.2t \quad , \end{aligned} \tag{6.39}$$

where t is the time index of the simulated volume. In Fig. 6.1b, the results of the order selection, again using the AIC order selection procedure with a penalty factor 3, of the simulated data are plotted. The parameters of the simulation were chosen to give approximately the same selection results, as can be seen by comparing Fig. 6.1a with Fig. 6.1b.

The most interesting parts of the histograms are those parts where the order of the trend model exceeds 1 or where the order of the AR model exceeds 4. In these parts, the orders are selected approximately equally often from the measured and simulated data. Therefore, we think that a model with a linear trend (polynomial order 1) and an AR(4) noise model gives a sufficiently accurate description of the data. A linear trend and AR models up to order 4 were therefore used in the simulation experiments of the next section. Note that we do not claim that the process underlying the data actually consists of a linear trend and AR(4) noise process, but only that there is not enough evidence to assume that higher order parameters are significantly present in the data. For AR orders lower than 4 or trend orders lower than 1 this could not be concluded since the histograms were always substantially different.

This analysis has been performed for several other data sets (results not shown). In these other data sets, the linear trend was generally present, but for some data sets AR(3) models turned out to give a sufficiently accurate description of the data. AR orders higher than 4 or trend orders higher than 1 were not needed to describe the data sets considered.

6.3.2 Simulation experiments

Simulation experiments were set up to detect brain activation. A simple block design activation scheme was used in which traces of 100 time points were generated with period equal to 20 (10 stimulus on, 10 stimulus off). This block stimulus was convolved with a standard HRF function [60] (`fmridesign` with default parameters and $TR = 2$) to get the activation pattern. Also, for each voxel a small linear trend increasing 0.1 per time point and a baseline of 100 were introduced. A linear trend model ($m = 1$) was used in the model, as well as the activation pattern. Note that when the trend (including baseline) is modeled correctly (as it is here), changing trend parameters does not influence the value of the likelihood function in its maximum, and thus the likelihood based test values are independent of the actual trend parameters. Several different noise models, based upon results of the previous section and selected to investigate different properties of the estimators, were used to generate fMRI data. These noise models were

$$\mathbf{AR}_a = [1 \quad -0.177 \quad -0.164 \quad -0.115 \quad -0.130] \tag{6.40}$$

$$\mathbf{AR}_b = [1 \quad -0.208 \quad -0.056 \quad 0.115] \tag{6.41}$$

$$\mathbf{AR}_c = [1 \quad -0.400 \quad 0.118 \quad 0.568] \tag{6.42}$$

The power spectral density (PSD) of these noise processes is plotted in Fig. 6.2a and the correlation functions in Fig. 6.2b. $\mathbf{AR}_a$ is a low frequency colored noise process. $\mathbf{AR}_b$ is almost white, but has slight excess power near one of the main frequencies present in the stimulus used. $\mathbf{AR}_c$ is stronger colored, also with the maximum power near one of the main frequencies present in the stimulus used. The simulations of the null-data (i.e. data containing no activation) with model $\mathbf{AR}_a$ used 20 000 independent traces, the simulations with model $\mathbf{AR}_b$ used 100 000 independent traces, and the simulations with model $\mathbf{AR}_c$ used 100 000 and 40 000 independent traces for the lengths 100 and 2500, respectively. To investigate the effect of changing SNR in the simulation experiments, the amplitude of the activation pattern was changed from 0 till 1.2, while the noise standard deviation was fixed to 1, which are realistic values of SNR in fMRI [**?**]. At each activation level, the time courses of 1000 voxels were generated. In this simulation experiment, the null hypothesis is given by Eq. (6.28), with $\boldsymbol{c} = \mathbf{0}$ and $\boldsymbol{C} = [1,\ 0,\ \ldots,\ 0]$ (the first column of $\boldsymbol{X}$ contains the activation related regressor).

Null distribution

The observed distribution of the test statistics under H_0, the null distribution, was compared with the theoretically known asymptotic null distribution. This is important since the asymptotic distribution might be used to compute thresholds for a given false alarm rate P_f. The comparison of the distributions

was made by observing the actual P_f of null data as a function of the theoretical asymptotic P_f. To help visualizing this, Fig. 6.3 shows a slice with the (falsely) detected active voxels in a real fMRI null experiment.

Activation sensitivity of the test statistics

For a fixed P_f of 0.1%, the detection rates (P_d) of the different tests were compared. Since it was observed that the observed null distribution of the test statistics was not equal to the asymptotic distribution, a correction was needed in order to make a fair comparison of the different test statistics and models. Therefore, in all experiments where different P_d values were compared, the observed null distribution of the simulated null data was used to compute threshold values to obtain a specified P_f.

6.3.3 Results of the simulation experiments

Null distribution

The thresholds for detecting activation can be computed by using the theoretical asymptotic distribution of the test statistics under H_0. However, this does not necessarily lead to an accurate P_f for time series with a limited trace length. In Fig. 6.4, the observed P_f of the different tests, with a linear trend and AR(4) noise model, are plotted as a function of the theoretical asymptotic P_f. Model $\mathbf{AR}_a$ was used to generate the noise for Fig. 6.4a, Fig. 6.4c and Fig. 6.4d and model $\mathbf{AR}_b$ was used to generate the noise for Fig. 6.4b. The first thing to note from Fig. 6.4a and Fig. 6.4b is that the distribution of the test statistics has not reached the asymptotic distribution for 80 or 100 samples per time series.

It can be noted from these figures as well that the observed P_f is larger for the Wald test than for the LR test, which in turn is larger than the P_f of the Rao test statistic. The P_f of the GLM test is somewhere in between. For linear models the ordering of the Wald, LR and Rao tests statistics can be proven to be as observed here, (see [68], p. 231). Since the models used in this chapter are nonlinear, this ordering might be different. However, we did not observe this in any data analyzed. Usually, as is the case in the presented figures, the Rao test statistic approximates the asymptotic distribution most accurately, especially in the most relevant region of false alarm rates between 0.01 and 0.001. However, even in this interval, the Rao test statistic has an actual P_f that might differ from the asymptotic value by a factor larger than 2. So for data series with 80 to 100 time points the asymptotic distribution cannot be reliably used to determine the thresholds of the test statistics.

When the length of the data series is increased, the asymptotic distribution is approached much more accurately, as is demonstrated in Fig. 6.4c and Fig. 6.4d were the trace length was 500 and 2500, respectively.

The main contribution to the difference between the observed and asymptotic distributions is the finite length of the time series. However, changing the regression model or noise process influences the distribution of the test statistic slightly. Therefore, the observed distributions shown in Fig. 6.4 cannot be reliably used for all different regression models and noise sources.

When the order of the noise model is below the order needed to give an accurate description of the data, the null distribution of the test statistics deviates from the asymptotic distribution as well, as might be expected since the model of the data is incomplete. In fact, this deviation can easily be much larger than the deviation caused by short data series. A demonstration of this is shown in Fig. 6.5, where the observed P_f is much larger or smaller than the P_f set using the asymptotic distribution for the AR(0), AR(1) and AR(2) noise models.

Apart from simulation experiments, the P_f of the tests under concern were computed for experimental fMRI null data sets. The asymptotic theoretical distribution was used to obtain the thresholds for the tests with a theoretical P_f of 1%. Fig. 6.3 shows the voxels that are detected as active with this threshold. For this threshold, the observed P_f of the Rao test statistic (Fig. 6.3a) is 2.0 ($\pm$0.7) times the asymptotic theoretical P_f. As is clear from Fig. 6.3b, Fig. 6.3c, and Fig. 6.3d, the LR, GLM, and Wald tests were observed to have even higher false alarm rates of approximately 4.4%, 5.7%, and 6.4%, respectively. This clearly demonstrates the need for correction of the P_f to obtain reliable activation detection.

Activation sensitivity of the test statistics

In the second simulation experiment, the activation sensitivity of the test statistics is investigated. The results are plotted in Fig. 6.6. An upper limit to the detection rate is included in these plots. This upper limit is the theoretical detection rate for the case in which the noise generating AR process and the noise variance σ_e^2 are known. In this case, all evaluated test statistics are equivalent and equal to

$$T_t = \frac{\left(\boldsymbol{C}\widehat{\boldsymbol{\theta}} - \boldsymbol{c}\right)^T \left(\boldsymbol{C}\boldsymbol{W}\boldsymbol{C}^T\right)^{-1} \left(\boldsymbol{C}\widehat{\boldsymbol{\theta}} - \boldsymbol{c}\right)}{\sigma_e^2} = \frac{\widehat{\theta}_1^{\,2}}{\sigma_{\theta_1}^2}, \tag{6.43}$$

still with $\boldsymbol{c} = \boldsymbol{0}$, $j = 1$ and $\boldsymbol{C} = [1,\ 0,\ \ldots,\ 0]$. Note that $\sigma_{\theta_1}^2 = \sigma_e^2 W_{1,1}$ is the variance of $\widehat{\theta}_1$. When the noise process is known, the tests are optimally sensitive, $\hat{\theta}_1$ will be normally distributed with mean value a (denoting the activation level used in the simulation), and the test statistic Eq. (6.43) has a non-central chi-squared distribution with 1 degree of freedom and non-centrality parameter $\lambda = \left(\frac{a}{\sigma_{\theta_1}}\right)^2$. The threshold value t_t, which can be computed from

$Pr(T_t > t_t; H_0) = \alpha$, can be used to compute the detection probability at each activation level, $p_d = Pr(T_t > t_t; H_1)$

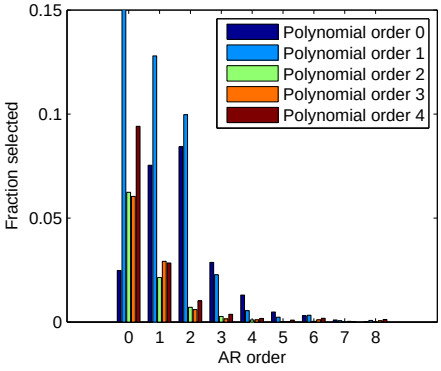
In practice, the coloring of the noise and the noise variance are not known. Therefore, this theoretical limit is unreachable. In Fig. 6.6a, it is visible that the Rao test statistic generally is the least sensitive to activation. The other three test statistics, LR, Wald, and GLM have approximately equal detection rates, although, by evaluating many simulations, it turns out that the LR test often has a slightly higher detection rate. However, it is far more important to use the correct noise model, as can be seen in Fig. 6.6b and Fig. 6.6c. These figures contain the results of the LR test statistic. The other statistics are almost overlapping and are therefore not plotted. When no or little color is present in the noise, as is the case with noise process $\mathbf{AR}_b$, the optimal detection rate can be reached by an AR(0) model (Fig. 6.6b). This is expected, since this is the model that can describe the data accurately with the lowest number of parameters. However, when the coloring of the data is stronger, as it is in noise model $\mathbf{AR}_c$, which is used for Fig. 6.6c, the reduced precision due to the extra parameters of the AR models is more than compensated for by an increase in accuracy of the model and thus, for a given activation amplitude, the detection rates of the AR(2) and AR(4) models are higher. So for (nearly) white noise processes, using a high order AR model (AR(4)) results in a "modest" performance loss, but, using a low order AR model (AR(0)) when there is strongly colored noise results in a "large" performance loss. This suggests that the order of the AR model should not be chosen too low.

6.4 Conclusions

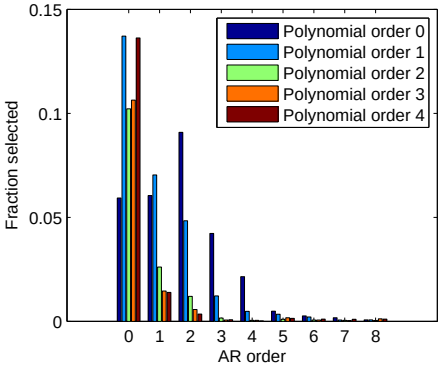
In this chapter, likelihood based tests for the detection of functional brain activity were presented. In contrast to the general linear model (GLM) tests, the proposed likelihood ratio tests allow direct incorporation of colored noise and do not require an explicit prewhitening step. Simulation results showed that the detection rate of the proposed likelihood ratio test is slightly, but significantly improved compared to the detection rate of the currently popular GLM based tests. Furthermore, it was demonstrated that thresholds based on theoretical, asymptotically valid null distributions of test statistics cannot be reliably used when the data series does not have more than a few hundred time points per voxel. In that case, thresholds obtained from observed null distributions should be used instead. Finally, it was shown that undermodeling of the (correlation structure of the) noise leads to inferior test results.

6.A Derivatives of the concentrated likelihood function

The concentrated likelihood function, as given in Eq. (6.27) is maximized to obtain the maximum likelihood function. However, in order to be able to

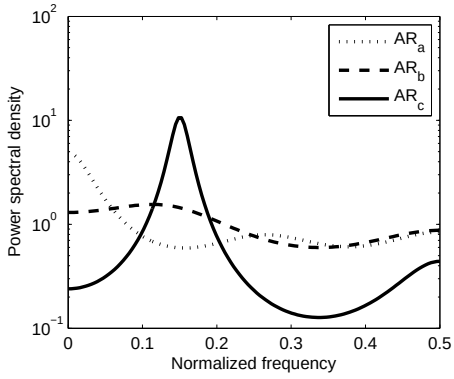


(a) Measured fMRI data

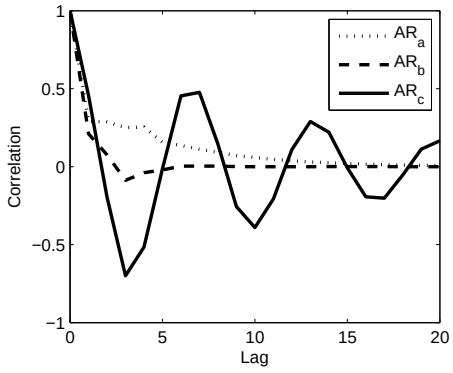


(b) Simulated fMRI data

Figure 6.1: *Histogram of the selected AR and polynomial orders from measured as well as simulated data, using the AIC criterion with penalty factor 3. See Eq. (6.39) for the parameters of the simulation model. 10 000 traces inside the brain were used for the measurements and 10 000 generated independent traces were used for the simulation. Note that for AR orders 5-8 the selection frequency for the measured and simulated data is approximately equal.*

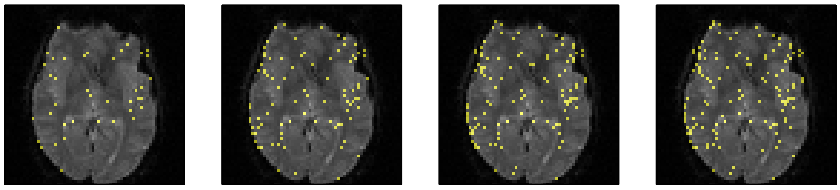


(a)



(b)

Figure 6.2: *(a) Power spectral density of the noise processes used in the simulations as function of the normalized frequency. (b) Correlation functions of the noise processes used in the simulations.*



(a) Test: Rao (b) Test: LR (c) Test: GLM (d) Test: Wald

Figure 6.3: *In these figures, the yellow dots are the voxels that were detected as active in real fMRI null data, by each of the four different test statistics. No stimulus was provided and thus the brain is expected to work in default mode in which the resting state networks are activated. The spontaneous activations in these networks, which are regarded as colored noise in our framework, are not correlated with the tested task and thus should not increase the false alarm rate. The threshold was computed to have a theoretical asymptotic P_f of 1% for all test statistics. An AR(4) model was used for the noise model, and the fMRI time series had a length of 80 points. Note the difference in false alarm rate of the different test statistics. For this asymptotic P_f, even the Rao test has an observed P_f that is 2.0 ($\pm$0.7) times higher than the theoretical value.*

efficiently maximize this likelihood function, the first and second derivatives are needed. In this section these derivatives of the logarithm of the concentrated likelihood function are derived. First, recall Eq. (6.27):

$$\ln\left(p(\boldsymbol{\Theta},\boldsymbol{\alpha}|\boldsymbol{Y})\right) = -\frac{nM}{2} - \frac{nM}{2}\ln\left(\frac{2\pi}{nM}\sum_{i=1}^{M}\widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})\right) - \frac{M}{2}\ln\left(|\boldsymbol{V}_r|\right) \tag{6.44}$$

with

$$\begin{aligned}\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}) = &\left(\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta}\right)^T \boldsymbol{V}_r^{-1}\left(\boldsymbol{y}_r - \boldsymbol{X}_{1:r}\boldsymbol{\theta}\right) + \\ &\sum_{t=r+1}^{n}\left(y_t - \boldsymbol{x}_t\boldsymbol{\theta} + \alpha_1(y_{t-1} - \boldsymbol{x}_{t-1}\boldsymbol{\theta}) + \ldots + \alpha_r(y_{t-r} - \boldsymbol{x}_{t-r}\boldsymbol{\theta})\right)^2.\end{aligned} \tag{6.45}$$

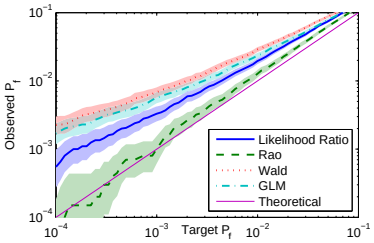
With the definitions:

$$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta} \tag{6.46}$$

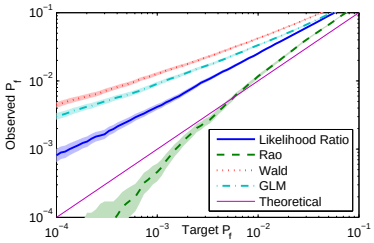
$$\boldsymbol{\varepsilon}_r = [\varepsilon_1,\ \ldots,\ \varepsilon_r]^T \tag{6.47}$$

$$\boldsymbol{\varepsilon}_t = [\varepsilon_t,\ \varepsilon_{t-1},\ \ldots,\ \varepsilon_{t-r}]^T \tag{6.48}$$

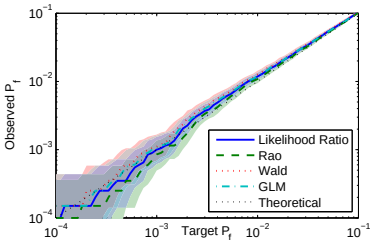
$$\boldsymbol{a} = [1,\ \alpha_1,\ \ldots,\ \alpha_r], \tag{6.49}$$



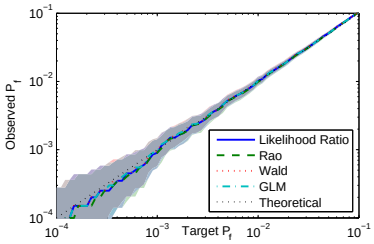
(a) Noise process $\mathbf{AR}_a$, trace length: 100



(b) Noise process $\mathbf{AR}_b$, trace length: 80



(c) Noise process $\mathbf{AR}_a$, trace length: 500



(d) Noise process $\mathbf{AR}_a$, trace length: 2500

Figure 6.4: *Logarithmic plot of P_f as a function of the P_f computed from the asymptotic distribution for the different test statistics. The diagonal corresponds with the asymptotic distribution. The shaded areas indicate the 95% confidence regions of the observed P_f as computed from the binomial counting statistics. The uncertainty indicated by these regions is caused by the finite number of voxel time series used in the simulations. For each simulation the noise contaminated time trace was modelled by a linear trend and an AR(4) model.*

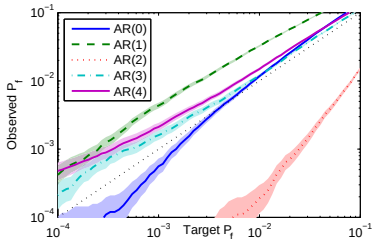
Eq. (6.45) can be simplified to:

$$\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha}) = \boldsymbol{\varepsilon}_r^T \boldsymbol{V}_r^{-1} \boldsymbol{\varepsilon}_r + \sum_{t=r+1}^{n} (a\boldsymbol{\varepsilon}_t)^2 \tag{6.50}$$

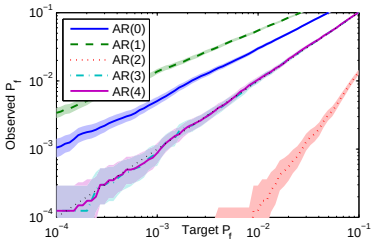
The first and second order derivatives of Eq. (6.27) with respect to $\boldsymbol{\Theta}$ and $\boldsymbol{\alpha}$ are given by

$$\frac{\partial \ln\left(p(\boldsymbol{\Theta},\boldsymbol{\alpha}|\boldsymbol{Y})\right)}{\partial \boldsymbol{\Theta}} = -\frac{nM}{2} \frac{\sum_{i=1}^{M} \frac{\partial \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}{\partial \boldsymbol{\Theta}}}{\sum_{i=1}^{M} \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})} - \frac{M}{2} \frac{\frac{\partial |\boldsymbol{V}_r|}{\partial \boldsymbol{\Theta}}}{|\boldsymbol{V}_r|} \tag{6.51}$$

$$\frac{\partial \ln\left(p(\boldsymbol{\Theta},\boldsymbol{\alpha}|\boldsymbol{Y})\right)}{\partial \boldsymbol{\alpha}} = -\frac{nM}{2} \frac{\sum_{i=1}^{M} \frac{\partial \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}}{\sum_{i=1}^{M} \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})} - \frac{M}{2} \frac{\frac{\partial |\boldsymbol{V}_r|}{\partial \boldsymbol{\alpha}}}{|\boldsymbol{V}_r|} \tag{6.52}$$



(a) Noise process $\mathbf{AR}_c$, 100 volumes



(b) Noise process $\mathbf{AR}_c$, 2500 volumes

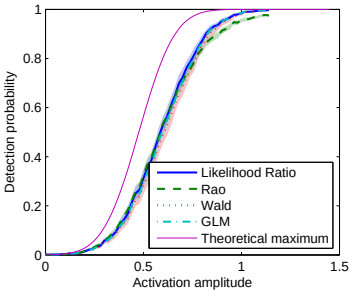
Figure 6.5: *This figure shows the observed distributions of the LR test statistic for different AR orders in the model. The axes in this figure are the same as in Fig. 6.4. Note that even for long time series (Figure (b)) the asymptotic distribution is not reached for the AR(1) and AR(2) models, since the noise is generated by an AR(3) process. Also note that for short time series none of the tests reaches the asymptotic χ^2_1 distribution.*

$$\frac{\partial^2 \ln\left(p(\boldsymbol{\Theta},\boldsymbol{\alpha}|\boldsymbol{Y})\right)}{\partial \boldsymbol{\Theta}^2} = \frac{nM}{2}\left(\frac{\sum_{i=1}^{M}\frac{\partial \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}{\partial \boldsymbol{\Theta}}}{\sum_{i=1}^{M}\widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}\right)^2 - \frac{nM}{2}\frac{\sum_{i=1}^{M}\frac{\partial^2 \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}{\partial \boldsymbol{\Theta}^2}}{\sum_{i=1}^{M}\widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})} + \left(\frac{M}{2}\frac{\frac{\partial |\boldsymbol{V}_r|}{\partial \boldsymbol{\Theta}}}{|\boldsymbol{V}_r|}\right)^2 - \frac{M}{2}\frac{\frac{\partial^2 |\boldsymbol{V}_r|}{\partial \boldsymbol{\Theta}^2}}{|\boldsymbol{V}_r|} \tag{6.53}$$

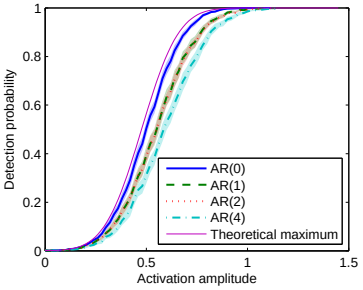
$$\frac{\partial^2 \ln\left(p(\boldsymbol{\Theta},\boldsymbol{\alpha}|\boldsymbol{Y})\right)}{\partial \boldsymbol{\Theta}\partial \boldsymbol{\alpha}} = \frac{nM}{2}\frac{\left(\sum_{i=1}^{M}\frac{\partial \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}{\partial \boldsymbol{\Theta}}\right)\left(\sum_{i=1}^{M}\frac{\partial \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right)}{\left(\sum_{i=1}^{M}\widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})\right)^2} - \frac{nM}{2}\frac{\sum_{i=1}^{M}\frac{\partial^2 \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}{\partial \boldsymbol{\Theta}\partial \boldsymbol{\alpha}}}{\sum_{i=1}^{M}\widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})} + \frac{M}{2}\frac{\left(\frac{\partial |\boldsymbol{V}_r|}{\partial \boldsymbol{\Theta}}\right)\left(\frac{\partial |\boldsymbol{V}_r|}{\partial \boldsymbol{\alpha}}\right)}{|\boldsymbol{V}_r|^2} - \frac{M}{2}\frac{\frac{\partial^2 |\boldsymbol{V}_r|}{\partial \boldsymbol{\Theta}^2}}{|\boldsymbol{V}_r|} \tag{6.54}$$

$$\frac{\partial^2 \ln\left(p(\boldsymbol{\Theta},\boldsymbol{\alpha}|\boldsymbol{Y})\right)}{\partial \boldsymbol{\alpha}^2} = \frac{nM}{2}\left(\frac{\sum_{i=1}^{M}\frac{\partial \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}}{\sum_{i=1}^{M}\widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}\right)^2 - \frac{nM}{2}\frac{\sum_{i=1}^{M}\frac{\partial^2 \widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^2}}{\sum_{i=1}^{M}\widetilde{Q}(\boldsymbol{Y}_i|\boldsymbol{\Theta}_i,\boldsymbol{\alpha})} + \left(\frac{M}{2}\frac{\frac{\partial |\boldsymbol{V}_r|}{\partial \boldsymbol{\alpha}}}{|\boldsymbol{V}_r|}\right)^2 - \frac{M}{2}\frac{\frac{\partial^2 |\boldsymbol{V}_r|}{\partial \boldsymbol{\alpha}^2}}{|\boldsymbol{V}_r|} \tag{6.55}$$

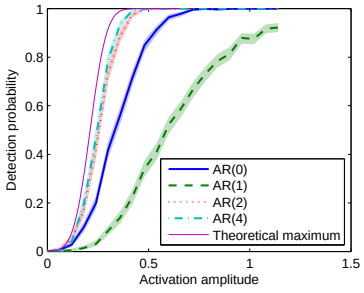
In order to evaluate these derivatives, the first and second derivatives of $\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})$ and $|\boldsymbol{V}_r|$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are needed. As the expansion of each of these



(a) Noise process $\mathbf{AR}_b$, modelled with AR(4) noise model



(b) Noise process $\mathbf{AR}_b$, LR test statistic



(c) Noise process $\mathbf{AR}_c$, LR test statistic

Figure 6.6: *Activation sensitivity of the different test statistics, with corrected false positive level. The theoretical maximum is reached when the true noise model and the noise variance are known. The number of timepoints is* 80 *for the figures (a) and (b) and* 100 *for figure (c). Figure (a) shows that the Rao test is less sensitive in detecting activation and that the LR test is slightly more sensitive than the others. Figures (b) and (c) show that the order of the noise model should be chosen carefully so as to avoid inferior activation sensitivity.*

terms is requires a few lines, each is expanded separately. In all the expansions, matrix identities, which were collected in the Matrix Cookbook [**?**], have been extensively used.

6.A.1 Derivatives of $|\boldsymbol{V}_r|$

This subsection expands the derivatives of $|\boldsymbol{V}_r|$. Since $\boldsymbol{V}_r$ does not depend on the parameters $\boldsymbol{\theta}$, all derivatives of $|\boldsymbol{V}_r|$ with respect to $\boldsymbol{\theta}$ are zero. The derivatives of $|\boldsymbol{V}_r|$ with respect to $\boldsymbol{\alpha}$ are not as trivial:

$$\frac{\partial|\boldsymbol{V}_r|}{\partial\boldsymbol{\alpha}} = |\boldsymbol{V}_r| \operatorname{tr}\left(\boldsymbol{V}_r^{-1}\frac{\partial\boldsymbol{V}_r}{\partial\boldsymbol{\alpha}}\right) \tag{6.56}$$

$$\begin{aligned}\frac{\partial^2|\boldsymbol{V}_r|}{\partial\alpha_i\partial\alpha_j} =& |\boldsymbol{V}_r|\left(\operatorname{tr}\left(\boldsymbol{V}_r^{-1}\frac{\partial^2\boldsymbol{V}_r}{\partial\alpha_i\partial\alpha_j}\right) + \operatorname{tr}\left(\boldsymbol{V}_r^{-1}\frac{\partial\boldsymbol{V}_r}{\partial\alpha_i}\right)\operatorname{tr}\left(\boldsymbol{V}_r^{-1}\frac{\partial\boldsymbol{V}_r}{\partial\alpha_j}\right)\right.\\ &\left. - \operatorname{tr}\left(\left(\boldsymbol{V}_r^{-1}\frac{\partial\boldsymbol{V}_r}{\partial\alpha_i}\right)\left(\boldsymbol{V}_r^{-1}\frac{\partial\boldsymbol{V}_r}{\partial\alpha_j}\right)\right)\right).\end{aligned} \tag{6.57}$$

Recall that $\boldsymbol{V}_r = g\,\text{toeplitz}(\boldsymbol{\rho}) = g\boldsymbol{R}$, where $g = \frac{\sigma_v^2}{\sigma_e^2}$ is the gain and $\boldsymbol{R} = \text{toeplitz}(\boldsymbol{\rho})$ is a function of $\boldsymbol{\rho} = [\rho(1),\ \ldots,\ \rho(r)]^T$ which is a function of $\boldsymbol{\alpha}$. Therefore, we can write

$$\frac{\partial\boldsymbol{V}_r}{\partial\boldsymbol{\alpha}} = g\frac{\partial\boldsymbol{R}}{\partial\boldsymbol{\rho}}\frac{\partial\boldsymbol{\rho}}{\partial\boldsymbol{\alpha}} + \frac{\partial g}{\partial\boldsymbol{\alpha}}\boldsymbol{R} \tag{6.58}$$

$$\frac{\partial^2\boldsymbol{V}_r}{\partial\boldsymbol{\alpha}^2} = 2\frac{\partial g}{\partial\boldsymbol{\alpha}}\frac{\partial\boldsymbol{R}}{\partial\boldsymbol{\rho}}\frac{\partial\boldsymbol{\rho}}{\partial\boldsymbol{\alpha}} + \frac{\partial^2 g}{\partial\boldsymbol{\alpha}^2}\boldsymbol{R} + g\frac{\partial\boldsymbol{R}}{\partial\boldsymbol{\rho}}\frac{\partial^2\boldsymbol{\rho}}{\partial\boldsymbol{\alpha}^2} + g\frac{\partial^2\boldsymbol{R}}{\partial\boldsymbol{\rho}^2}\frac{\partial\boldsymbol{\rho}}{\partial\boldsymbol{\alpha}}\frac{\partial\boldsymbol{\rho}}{\partial\boldsymbol{\alpha}} \tag{6.59}$$

Since $\boldsymbol{R} = \text{toeplitz}(\boldsymbol{\rho})$ we can compute $\frac{\partial\boldsymbol{R}}{\partial\boldsymbol{\rho}}$,

$$\frac{\partial(\boldsymbol{R})_{kl}}{\partial\rho_i} = \quad \delta_{k,l+i} + \delta_{k,l-i} \quad \forall \quad i,k,l \in \{1\ldots r\}\ , \tag{6.60}$$

with the $\delta_{i,j}$ Kronecker delta function, so the second order order derivative of $\boldsymbol{R}$ is zero.

The vector with correlation coefficients $\boldsymbol{\rho}$ can be computed from $\boldsymbol{\alpha}$ with the Yule-Walker equations

$$M(\boldsymbol{\alpha})\boldsymbol{\rho} = -\boldsymbol{\alpha} \quad \Rightarrow \quad \boldsymbol{\rho} = -M(\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha} \tag{6.61}$$

where

$$M_{ij}(\boldsymbol{\alpha}) = \delta_{ij} + \alpha_{i+j} + \alpha_{i-j} \tag{6.62}$$

and $\alpha_i = 0$ for $i \leq 0$ or $i > r$. This allows to immediately compute the derivative of $\boldsymbol{M}$ to α:

$$\frac{\partial M_{ij}}{\partial\alpha_k} = \frac{\partial}{\partial\alpha_k}\left(\delta_{ij} + \alpha_{i+j} + \alpha_{i-j}\right) \tag{6.63}$$

$$= \delta_{i+j,k} + \delta_{i-j,k}, \tag{6.64}$$

and the second derivative is (obviously) zero. The derivatives of $\boldsymbol{\rho}$ with respect to $\boldsymbol{\alpha}$ are given by

$$\frac{\partial \boldsymbol{\rho}}{\partial \alpha_i} = -\left(M(\boldsymbol{\alpha})^{-1}\right)_{:,i} - \frac{\partial M(\boldsymbol{\alpha})^{-1}}{\partial \alpha_i}\boldsymbol{\alpha} \tag{6.65}$$

$$\frac{\partial^2 \boldsymbol{\rho}}{\partial \alpha_i \partial \alpha_j} = -\left(\frac{\partial M(\boldsymbol{\alpha})^{-1}}{\partial \alpha_j}\right)_{:,i} - \frac{\partial^2 M(\boldsymbol{\alpha})^{-1}}{\partial \alpha_i \partial \alpha_j}\boldsymbol{\alpha} - \left(\frac{\partial M(\boldsymbol{\alpha})^{-1}}{\partial \alpha_i}\right)_{:,j} \tag{6.66}$$

where an index $\boldsymbol{M}_{i,:}$ indicates the i^{th} row of the matrix $\boldsymbol{M}$, the index $\boldsymbol{M}_{:,i}$ indicates the i^{th} column of the matrix $\boldsymbol{M}$ and

$$\frac{\partial \boldsymbol{M}(\boldsymbol{\alpha})^{-1}}{\partial \alpha_k} = -\sum_{i,j=1}^{r} \left(\boldsymbol{M}(\boldsymbol{\alpha})^{-1}\right)_{:,i} \left(\delta_{i+j,k} + \delta_{i-j,k}\right) \left(\boldsymbol{M}(\boldsymbol{\alpha})^{-1}\right)_{j,:} \tag{6.67}$$

$$\begin{aligned}\frac{\partial^2 \boldsymbol{M}(\boldsymbol{\alpha})^{-1}}{\partial \alpha_k \partial \alpha_l} =& 2 \sum_{i,j,m,n=1}^{r} \left(\boldsymbol{M}(\boldsymbol{\alpha})^{-1}\right)_{:,m} \left(\delta_{m+n,k} + \delta_{m-n,k}\right) \left(\boldsymbol{M}(\boldsymbol{\alpha})^{-1}\right)_{n,i} \times \\ & \left(\delta_{i+j,k} + \delta_{i-j,k}\right) \left(\boldsymbol{M}(\boldsymbol{\alpha})^{-1}\right)_{j,:}\end{aligned} \tag{6.68}$$

The gain g can be computed from reflection coefficients rc or with $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ as in Eq. (6.6)

$$g = \frac{1}{1 - \sum_{k=1}^{r} \alpha_k \rho_k} = \frac{1}{\prod_{k=1}^{r}(1 - rc_k^2)} \tag{6.69}$$

Reflection coefficients rc might be used for faster computing and to be able to more easily include stability constraint on the AR model (i.e. finite gain), since a AR model is stable *iff* $|rc_i| \leq 1 \quad \forall \quad i = 1 \dots r$. The derivative of g with respect to $\boldsymbol{\alpha}$ is given by

$$\begin{aligned}\frac{\partial g}{\partial \alpha_i} =& \frac{\partial}{\partial \alpha_i} \frac{1}{1 + \boldsymbol{\alpha}^T \boldsymbol{M}(\boldsymbol{\alpha})^{-1} \boldsymbol{\alpha}} \\ =& -\frac{\left(\boldsymbol{M}(\boldsymbol{\alpha})^{-1}\right)_{i,:}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T \left(\boldsymbol{M}(\boldsymbol{\alpha})^{-1}\right)_{:,i} + \boldsymbol{\alpha}^T \frac{\partial \boldsymbol{M}(\boldsymbol{\alpha})^{-1}}{\partial \alpha_i}\boldsymbol{\alpha}}{\left(1 + \boldsymbol{\alpha}^T \boldsymbol{M}(\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}\right)^2}\end{aligned} \tag{6.70}$$

$$\frac{\partial^2 g}{\partial \alpha_i \partial \alpha_j} = 2\frac{\left(\left(\boldsymbol{M}(\boldsymbol{\alpha})^{-1}\right)_{i,:}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T \left(\boldsymbol{M}(\boldsymbol{\alpha})^{-1}\right)_{:,i} + \boldsymbol{\alpha}^T \frac{\partial \boldsymbol{M}(\boldsymbol{\alpha})^{-1}}{\partial \alpha_i}\boldsymbol{\alpha}\right)^2}{\left(1 + \boldsymbol{\alpha}^T \boldsymbol{M}(\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}\right)^3} \tag{6.71}$$

$$- \frac{\begin{array}{c}\left(M(\boldsymbol{\alpha})^{-1}\right)_{i,j} + \left(M(\boldsymbol{\alpha})^{-1}\right)_{j,i} + \frac{\partial \boldsymbol{M}(\boldsymbol{\alpha})^{-1})_{i,:}}{\partial \alpha_j}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T \frac{\partial (\boldsymbol{M}(\boldsymbol{\alpha})^{-1})_{:,i}}{\partial \alpha_j} \dots \\ \dots + \alpha_j \left(\frac{\partial \boldsymbol{M}(\boldsymbol{\alpha})^{-1}}{\partial \alpha_i}\right)_{j,:} \boldsymbol{\alpha} + \boldsymbol{\alpha}^T \frac{\partial^2 \boldsymbol{M}(\boldsymbol{\alpha})^{-1}}{\partial \alpha_j \partial \alpha_i}\boldsymbol{\alpha} + \boldsymbol{\alpha}^T \left(\frac{\partial \boldsymbol{M}(\boldsymbol{\alpha})^{-1}}{\partial \alpha_i}\right)_{:,j} \alpha_j\end{array}}{\left(1 + \boldsymbol{\alpha}^T \boldsymbol{M}(\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}\right)^2}$$

6.A.2 Derivatives of $\widetilde{Q}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha})$

This subsection describes the derivatives of $\widetilde{Q}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\alpha})$. As this term depends both on $\boldsymbol{\alpha}$ and θ, both derivatives have to be computed.

First the derivative of $\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})$ with respect to $\boldsymbol{\theta}$:

$$\frac{\partial \widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \theta_i} = -2\boldsymbol{\varepsilon}_r^T \boldsymbol{V}_r^{-1} \boldsymbol{X}_{r,i} - \sum_{t=r+1}^{n} 2\boldsymbol{\varepsilon}_t^T \boldsymbol{a}\boldsymbol{a}^T \boldsymbol{X}_{t,i} \tag{6.72}$$

$$\frac{\partial^2 \widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \theta_i \partial \theta_j} = 2\boldsymbol{X}_{r,i}^T \boldsymbol{V}_r^{-1} \boldsymbol{X}_{r,j} + \sum_{t=r+1}^{n} 2\boldsymbol{X}_{t,i}^T \boldsymbol{a}\boldsymbol{a}^T \boldsymbol{X}_{t,j} \tag{6.73}$$

were we noted that V_r is symmetric, $\boldsymbol{\varepsilon}$ is the only term that depends on $\boldsymbol{\theta}$ and

$$\frac{\partial \boldsymbol{\varepsilon}}{\partial \theta_i} = -\boldsymbol{X}_i \tag{6.74}$$

$$\frac{\partial \boldsymbol{\varepsilon}_r}{\partial \theta_i} = -\boldsymbol{X}_{r,i} \tag{6.75}$$

$$\frac{\partial \boldsymbol{\varepsilon}_t}{\partial \theta_i} = -\boldsymbol{X}_{t,i} \tag{6.76}$$

where $\boldsymbol{X}_i$ is the i^{th} column of the matrix $\boldsymbol{X}$ and

$$\boldsymbol{X}_{r,i} = [X_{1,i},\ \ldots,\ X_{r,i}]^T \tag{6.77}$$

$$\boldsymbol{X}_{t,i} = [X_{t,i},\ X_{t-1,i},\ \ldots,\ X_{t-r,i}]^T, \tag{6.78}$$

corresponding to Eq. (6.47) and Eq. (6.48). Next the derivatives of $\widetilde{Q}(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$:

$$\frac{\partial \widetilde{Q}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \alpha_i} = -\boldsymbol{\varepsilon}_r^T \boldsymbol{V}_r^{-1} \frac{\partial \boldsymbol{V}_r}{\partial \alpha_i} \boldsymbol{V}_r^{-1} \boldsymbol{\varepsilon}_r + \sum_{t=r+1}^{n} 2\varepsilon_{t-i} \left(\boldsymbol{a}^T \boldsymbol{\varepsilon}_t\right) \tag{6.79}$$

$$\begin{aligned}\frac{\partial^2 \widetilde{Q}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \alpha_i \partial \alpha_j} = -\boldsymbol{\varepsilon}_r^T \Bigg(&-\boldsymbol{V}_r^{-1} \frac{\partial \boldsymbol{V}_r}{\partial \alpha_j} \boldsymbol{V}_r^{-1} \frac{\partial \boldsymbol{V}_r}{\partial \alpha_i} \boldsymbol{V}_r^{-1} + \boldsymbol{V}_r^{-1} \frac{\partial^2 \boldsymbol{V}_r}{\partial \alpha_i \partial \alpha_j} \boldsymbol{V}_r^{-1} \\ &-\boldsymbol{V}_r^{-1} \frac{\partial \boldsymbol{V}_r}{\partial \alpha_i} \boldsymbol{V}_r^{-1} \frac{\partial \boldsymbol{V}_r}{\partial \alpha_j} \boldsymbol{V}_r^{-1} \Bigg) \boldsymbol{\varepsilon}_r + \sum_{t=r+1}^{n} 2\varepsilon_{t-i}\varepsilon_{t-j}\end{aligned} \tag{6.80}$$

where we note that $\frac{\partial \boldsymbol{V}_r}{\partial \alpha_i}$ and $\frac{\partial^2 \boldsymbol{V}_r}{\partial \alpha_i \alpha_j}$ have been derived in the previous subsection.

The final second derivative is given by:

$$\frac{\partial^2 \widetilde{Q}(\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \theta_i \partial \alpha_j} = 2\boldsymbol{\varepsilon}_r^T \boldsymbol{V}_r^{-1} \frac{\partial \boldsymbol{V}_r}{\partial \alpha_j} \boldsymbol{V}_r^{-1} \boldsymbol{X}_{r,i} + \sum_{t=r+1}^{n} \left(2\varepsilon_{t-j} \boldsymbol{a}^T \boldsymbol{X}_{t,i} + 2X_{t-j,i} \boldsymbol{a}^T \boldsymbol{\varepsilon}_t\right) \tag{6.81}$$

6.B The Fisher score vector and the Fisher information matrix

For the Rao and Wald tests the Fisher information matrix is needed. Therefore, the Fisher score vector and Fisher information matrix are computed in this appendix.

The Fisher score vector of the data set $\boldsymbol{y}$ with respect to the parameters $\boldsymbol{\tau} = (\boldsymbol{\theta}^T \boldsymbol{\alpha}^T \sigma_e^2)^T$ is defined as the $(m + r + 1) \times 1$ vector

$$s(\boldsymbol{\tau}) = \frac{\partial \ln p}{\partial \boldsymbol{\tau}}, \tag{6.82}$$

with p the joint PDF of the observations described by Eq. (6.17). It can be shown that the expectation of the Fisher score (evaluated at the true values of the parameters) is equal to zero [74], that is,

$$\mathbb{E}\left[s(\boldsymbol{\tau})\right] = \mathbf{0}, \tag{6.83}$$

with $\mathbf{0}$ the $(m+r+1) \times 1$ null vector. The $(m+r+1) \times (m+r+1)$ covariance matrix of the Fisher score is therefore given by [74]

$$\boldsymbol{F} = \mathbb{E}\left[s(\boldsymbol{\tau}) s^T(\boldsymbol{\tau})\right] = \mathbb{E}\left[\frac{\partial \ln p}{\partial \boldsymbol{\tau}} \left(\frac{\partial \ln p}{\partial \boldsymbol{\tau}}\right)^T\right]. \tag{6.84}$$

This covariance matrix is called the Fisher information matrix [74]. It can be shown that under certain regularity conditions $\boldsymbol{F}$ may alternatively be written as

$$\boldsymbol{F} = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\tau}^2}\right]. \tag{6.85}$$

The Fisher matrix may be written in the form

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\alpha}} & \boldsymbol{F}_{\boldsymbol{\theta}\sigma_e^2} \\ \boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\theta}} & \boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & \boldsymbol{F}_{\boldsymbol{\alpha}\sigma_e^2} \\ \boldsymbol{F}_{\sigma_e^2\boldsymbol{\theta}} & \boldsymbol{F}_{\sigma_e^2\boldsymbol{\alpha}} & \boldsymbol{F}_{\sigma_e^2\sigma_e^2} \end{bmatrix}, \tag{6.86}$$

where

$$\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\theta}^2}\right] \qquad \boldsymbol{F}_{\boldsymbol{\theta}\sigma_e^2} = \boldsymbol{F}_{\sigma_e^2\boldsymbol{\theta}}^T = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\theta} \partial \sigma_e^2}\right]$$

$$\boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\alpha}^2}\right]$$

$$\boldsymbol{F}_{\boldsymbol{\alpha}\sigma_e^2} = \boldsymbol{F}_{\sigma_e^2\boldsymbol{\alpha}}^T = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\alpha} \partial \sigma_e^2}\right]$$

$$\boldsymbol{F}_{\sigma_e^2\sigma_e^2} = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial (\sigma_e^2)^2}\right]$$

$$\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\alpha}} = \boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\theta}}^T = -\mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \boldsymbol{\theta} \partial \boldsymbol{\alpha}}\right]$$

with $\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} \in \mathbb{R}^{m \times m}$,$\boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \in \mathbb{R}^{r \times r}$,$\boldsymbol{F}_{\sigma_e^2\sigma_e^2} \in \mathbb{R}^{1 \times 1}$,$\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\alpha}} \in \mathbb{R}^{m \times r}$, $\boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\theta}} \in \mathbb{R}^{r \times m}$,$\boldsymbol{F}_{\boldsymbol{\theta}\sigma_e^2} \in \mathbb{R}^{m \times 1}$. It can be shown that all elements of $\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\alpha}}$, $\boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\theta}}$, $\boldsymbol{F}_{\boldsymbol{\alpha}\sigma_e^2}$,

$\boldsymbol{F}_{\sigma_e^2\boldsymbol{\alpha}}$, $\boldsymbol{F}_{\boldsymbol{\theta}\sigma_e^2}$ and $\boldsymbol{F}_{\sigma_e^2\boldsymbol{\theta}}$ are equal to zero. This means that Eq. (6.86) simplifies to

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{F}_{\sigma_e^2\sigma_e^2} \end{bmatrix}. \tag{6.87}$$

It can be shown that

$$\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{1}{\sigma_e^2}\boldsymbol{X}_{1:r}^T\boldsymbol{V}_r^{-1}\boldsymbol{X}_{1:r} + \frac{1}{\sigma_e^2}\sum_{t=r+1}^{n}\boldsymbol{X}_{t:t-r}^T\boldsymbol{A}\boldsymbol{X}_{t:t-r}, \tag{6.88}$$

where

$$\boldsymbol{X}_{t:t-r} = \begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{x}_{t-1} \\ \vdots \\ \boldsymbol{x}_{t-r} \end{bmatrix} \tag{6.89}$$

and

$$\boldsymbol{A} = \begin{bmatrix} 1 & \boldsymbol{\alpha}^T \\ \boldsymbol{\alpha} & \boldsymbol{\alpha}\boldsymbol{\alpha}^T \end{bmatrix}. \tag{6.90}$$

For activation detection only $\boldsymbol{\theta}$ is used in the test statistics. Therefore, and because $\boldsymbol{F}$ is block diagonal, the only term of $\boldsymbol{F}$ that is needed is $\boldsymbol{F}_{\boldsymbol{\theta}\boldsymbol{\theta}}$.

Chapter 7

Optimal experimental design for Diffusion Kurtosis Imaging

The main part of the work described in this chapter is published as:

Dirk H. J. Poot, Arnold J. den Dekker, Eric Achten, Marleen Verhoye and Jan Sijbers

Optimal experimental design for Diffusion Kurtosis Imaging

in: IEEE Transactions on Medical Imaging, Volume 29, Issue 3, March 2010, Pages 819 - 829

Abstract

Diffusion Kurtosis Imaging (DKI) is a new magnetic resonance imaging model that describes the non-Gaussian diffusion behavior in tissues. It has recently been shown that DKI parameters, such as the radial or axial kurtosis, are more sensitive to brain physiology changes than the well known diffusion tensor imaging (DTI) parameters in several white and grey matter structures.

In order to estimate either DTI or DKI parameters with maximum precision, the diffusion weighting gradient settings that are applied during the acquisition need to be optimized. Indeed, it has been shown previously that optimizing the set of diffusion weighting gradient settings can have a significant effect on the precision with which DTI parameters can be estimated. In this chapter, we focus on the optimization of DKI gradients settings. Commonly, DKI data are acquired using a standard

set of diffusion weighting gradients with fixed directions and with regularly spaced gradient strengths. In this work, we show that such gradient settings are suboptimal with respect to the precision with which DKI parameters can be estimated. Furthermore, the gradient directions and the strengths of the diffusion weighted MR images are optimized by minimizing the Cramér Rao lower bound of DKI parameters. The impact of the optimized gradient settings is evaluated, both on simulated as well as experimentally recorded datasets. It is shown that the precision with which the kurtosis parameters can be estimated, increases substantially by optimizing the gradient settings.

keywords Diffusion Kurtosis Imaging, Diffusion weighted MRI, Experimental design, optimization, gradient settings

7.1 Introduction

Diffusion weighted (DW) magnetic resonance imaging (MRI) is the only method available that measures non-invasively diffusion properties of tissues. Knowledge of these diffusion properties allows the characterization of intrinsic features of tissue microdynamics and microstructure, such as cell permeability [69, 70]. The diffusion of water molecules within a voxel is characterized by a statistical distribution describing the random displacement of these molecules during a fixed-time diffusion process. A popular model to describe this distribution is the Diffusion Tensor (DT) model, which assumes the diffusion distribution to be Gaussian.

Previously, it has been reported that the diffusion distribution in the human brain is generally non-Gaussian [70–72], due to diffusion restriction by cell membranes and compartments of different sizes present in the neural tissue. Since the DTI model is limited to Gaussian diffusion only, the model can generally not describe these diffusion profiles accurately. Diffusion kurtosis imaging (DKI) was recently proposed as an extension to the Gaussian DT model. It was shown that DKI allows a better detection and characterization of changes in various white and grey matter structures [73]. In addition to the second central moment of the diffusion distribution, DKI also measures the kurtosis excess of that distribution. The kurtosis excess is defined as the fourth central moment of the distribution divided by the square of the variance minus 3 [74]. The "minus 3" term is often explained as a correction to make the kurtosis zero for a Gaussian distribution. Hence, compared to a DTI model, the inclusion of the kurtosis excess allows a more accurate description of the diffusion properties of neural tissues [71, 75].

Commonly, DKI data are acquired using a standard set of diffusion weighting gradients with fixed directions and with regularly spaced gradient strengths. As is shown in this chapter, such imaging settings are suboptimal with respect to the precision with which DKI parameters can be estimated from the DW-

MR images. This precision strongly depends on the directions and strengths (b-values) of the diffusion weighting gradients during the DW-MR acquisition. A lower bound on the variance (precision) of any unbiased estimator is given by the Cramér Rao Lower Bound (CRLB) [15]. In this chapter, the gradient settings of the DKI experiments are optimized by minimizing the CRLB of DKI parameters of interest.

Previously, several studies were published that optimized DW-MRI settings for estimating DTI parameters such as the fractional anisotropy (FA) and the mean diffusivity (MD) [76, 77]. In [76], for example, the CRLB of various diffusion tensor parameters was minimized with respect to the b-values of the DW-MR images. In this chapter, that work is extended to the optimization of DKI gradient settings. Furthermore, since the DKI model has significantly more parameters than the DTI model, a new numerical optimization strategy is developed. Extensive simulation experiments validated with real experiments show that using the optimized gradient settings allows estimation of DKI parameters with a substantially higher precision.

This chapter is organized as follows. Section 7.2 describes the DKI signal model, the DKI parameters, the CRLB for estimating these parameters, and the optimization method. Next, in Section 7.3, simulations and real experiments are presented which investigate the robustness and improvement of the performance of the optimized gradient settings, compared to the traditional gradient settings. Finally, in Section 7.5, the conclusions are drawn.

7.2 Methods

To obtain the most precise DKI parameter estimates, the directions and b-values of the diffusion weighting gradients need to be optimized. First, in Section 7.2.1, the kurtosis imaging model is explained. Next, Section 7.2.2 describes the computation of the CRLB for estimating kurtosis estimators. In Section 7.2.3, various kurtosis parameters are introduced. After that, Section 7.2.4 elaborates on the optimization of the gradient settings. Finally, Sections 7.2.5 describes an efficient optimization strategy.

7.2.1 Diffusion Kurtosis Imaging

The diffusion of hydrogen atoms in a voxel can be characterized by a 3D symmetric probability density function (PDF) $f(\boldsymbol{x}, t)$, where the random variable $\boldsymbol{x}$ denotes the random displacement of molecules during a diffusion process in a time interval t. It depends on the microstructure of the voxel, which is generally different for each voxel.

DW-MRI does not measure the diffusion PDF directly. The gradients that are applied during the diffusion weighting introduce a change in the phase of the precessing and diffusing hydrogen atoms, which leads to a reduction of the

magnitude of the DW image when compared to the unweighted image. The magnitude of a DW image depends on the diffusion in the direction of the applied diffusion weighting gradient, specified by $\boldsymbol{q}$. The q-space vector $\boldsymbol{q}$ is given by

$$\boldsymbol{q} = \gamma\delta\boldsymbol{G} \quad , \tag{7.1}$$

where γ [rad/s,T] is the gyromagnetic ratio, δ [s] is the duration of the pulsed gradients, and $\boldsymbol{G} = G\boldsymbol{g}$ is a vector of the magnitude and direction of the applied diffusion weighting gradient where G [T/m] specifies the strength and $\boldsymbol{g}$ is the unit length gradient direction vector. To take the duration of the diffusion gradient pulses into account, the b-value is usually defined as [6]

$$b = q^2 t \, [\mathrm{s/m^2}], \tag{7.2}$$

where $q = |\boldsymbol{q}|$ and $t = (\Delta - \delta/3)$ in which Δ [s] is the time separation between the leading edges of the diffusion gradient pulses [6].

Let $x = \boldsymbol{g}^T\boldsymbol{x}$ be the component of a displacement vector $\boldsymbol{x}$ in the direction of $\boldsymbol{g}$. The PDF of $\boldsymbol{x}$ in the direction of $\boldsymbol{g}$ is then given by:

$$f_{\boldsymbol{g}}(x;t) = \int_{\boldsymbol{g}^T\boldsymbol{x}=x} f(\boldsymbol{x};t)d\boldsymbol{x}. \tag{7.3}$$

That is, $f_{\boldsymbol{g}}(x;t)$ is $f(\boldsymbol{x};t)$ integrated over the 2 dimensions orthogonal to $\boldsymbol{g}$.

The diffusion coefficient $D_{\boldsymbol{g}}$ in the direction of the gradient $\boldsymbol{g}$, which is the variance of the diffusion in that direction, is given by:

$$D_{\boldsymbol{g}} = \frac{1}{2t}\mathbb{E}_{f_{\boldsymbol{g}}}[x^2], \tag{7.4}$$

where $\mathbb{E}_{f_{\boldsymbol{g}}}[.]$ is the expectation operator with respect to $f_{\boldsymbol{g}}$. The excess kurtosis $K_{\boldsymbol{g}}$ of the diffusion in the direction $\boldsymbol{g}$ is given by [74]

$$K_{\boldsymbol{g}} = \frac{\mathbb{E}_{f_{\boldsymbol{g}}}[x^4]}{\mathbb{E}_{f_{\boldsymbol{g}}}[x^2]^2} - 3. \tag{7.5}$$

The phase shift induced by the diffusion weighting gradients along the direction $\boldsymbol{g}$ is a linear function of x and q. Therefore, the magnitude of the noise free MR signal after diffusion weighting with the gradient $\boldsymbol{q}$, is given by

$$\mathcal{A}(\boldsymbol{q}) = A_0\mathbb{E}_{f_{\boldsymbol{g}}}[e^{iqx}] = A_0\int_{x=-\infty}^{\infty} e^{iqx} f_{\boldsymbol{g}}(x;t)dx, \tag{7.6}$$

where A_0 is the non diffusion weighted signal intensity. Note that Eq. (7.6) is equal to the characteristic function of $f_{\boldsymbol{g}}(x;t)$, multiplied by A_0.

An approximate parametric model of Eq. (7.6) can be derived from a Taylor series expansion around $q = 0$ of the natural logarithm of $\mathcal{A}(\boldsymbol{q})$ [75]:

$$\begin{aligned}\ln \mathcal{A}(\boldsymbol{q}) &= \ln A_0 - b \sum_{i,j=1}^{3} g_i g_j D_{ij} + \\ &\frac{b^2}{6} \left(\sum_{i=1}^{3} \frac{D_{ii}}{3} \right)^2 \sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l W_{ijkl} + O(q^6)\end{aligned} \tag{7.7}$$

in which g_i is the i^{th} component of $\boldsymbol{g}$, D_{ij} is the ij^{th} element of the 2^{nd} order diffusion tensor $\boldsymbol{D}$ and W_{ijkl} is the $ijkl^{\text{th}}$ element of the 4^{th} order kurtosis tensor $\boldsymbol{W}$. A detailed derivation of Eq. (7.7) can be found in Appendix 7.A. The elements D_{ij} and W_{ijkl} are defined as

$$D_{ij} = \frac{1}{2t} \mathbb{E}_f[x_i x_j] \tag{7.8}$$

and

$$W_{ijkl} = 9 \frac{\mathbb{E}_f[x_i x_j x_k x_l] - \mathbb{E}_f[x_i x_j]\mathbb{E}_f[x_k x_l] - \mathbb{E}_f[x_i x_k]\mathbb{E}_f[x_j x_l] - \mathbb{E}_f[x_i x_l]\mathbb{E}_f[x_j x_k]}{\left(\mathbb{E}_f[\boldsymbol{x}^T \boldsymbol{x}]\right)^2}, \tag{7.9}$$

respectively, where $\mathbb{E}_f[.]$ is the expectation operator with respect to f and with x_i the i^{th} component of $\boldsymbol{x}$ [75]. Note that both $\boldsymbol{D}$ and $\boldsymbol{W}$ are fully symmetric with respect to an interchange of indices.

From Eq. (7.7), the following approximate parametric DKI model of the magnitude of the noise free diffusion weighted MR signal can then be obtained:

$$A(\boldsymbol{q}; \boldsymbol{\theta}) = A_0 \exp\left(-b \sum_{i,j=1}^{3} g_i g_j D_{ij}^{\text{app}} + \frac{b^2}{6} \left(\sum_{i=1}^{3} \frac{D_{ii}^{\text{app}}}{3} \right)^2 \sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l W_{ijkl}^{\text{app}} \right), \tag{7.10}$$

where the diffusion and kurtosis tensors $\boldsymbol{D}$ and $\boldsymbol{W}$ are replaced by the apparent diffusion and kurtosis tensors $\boldsymbol{D}^{app}$ and $\boldsymbol{W}^{app}$, respectively [**?**]. It is known that, for short duration δ of the diffusion gradient pulse, the apparent diffusion and kurtosis tensors approach the true diffusion and kurtosis tensors given by Eq. (7.8) and Eq. (7.9), respectively [75]. In Eq. (7.10), $\boldsymbol{\theta}$ denotes a 22×1 parameter vector composed of the following scalar valued parameters: A_0, 6 parameters representing the independent elements of the symmetric tensor $\boldsymbol{D}^{\text{app}}$, and 15 parameters representing the independent elements of the fully symmetric tensor $\boldsymbol{W}^{\text{app}}$.

Remarks on the implementation

parameterization For the actual implementation of Eq. (7.10), the parameter vector $\boldsymbol{\theta}$ needs to be explicitly defined. In our diffusion kurtosis model we used a simple parameterization consisting of the independent elements of the Diffusion and Kurtosis tensor:

$$\boldsymbol{d} = [D_{ij;1\leq i\leq j\leq 3}]^T \qquad \text{6 element vector} \tag{7.11}$$

$$\boldsymbol{w} = [W_{ijkl;1\leq i\leq j\leq k\leq l\leq 3}]^T \qquad \text{15 element vector} \tag{7.12}$$

$$\boldsymbol{\theta} = [A_0 \quad \boldsymbol{d}^T \quad \boldsymbol{w}^T]^T. \tag{7.13}$$

Other parameterizations are possible and, potentially, this might influence the estimated values. For example, it is possible to use a parameterization of the diffusion tensor $D_{i,j}$ that guarantees positive definiteness of this tensor. A possible alternative for the parameterization of the kurtosis tensor is the selection of the coefficients of

$$M_{ijkl}^{\text{app}} = \left(\sum_{m=1}^{3} \frac{D_{mm}^{\text{app}}}{3}\right)^2 W_{ijkl}^{\text{app}}, \tag{7.14}$$

instead of W_{ijkl}, as this is more directly related to the 4^{th} derivative. As the different parameterizations describe the same function they will find the same optimal position. That is, unless constraints introduced in a specific parameterization, such as positive definiteness of the diffusion tensor $\boldsymbol{D}$, are violated in this optimum. However, the parameterization that is used might influence the rates of convergence and/or differ in computational complexity. So, the CPU time required to reach the optimum might differ for different parameterizations. Since there are $1 + 5 + 15 = 22$ parameters to be optimized for each voxel, the efficiency of the estimator is important. Therefore, in the optimization, it is best to choose the parameterization with which the estimation requires the least amount of CPU time. When desired, the parameters can be transformed to any different parameterization after the optimization.

derivatives To be able to estimate $\boldsymbol{\theta}$ with the ML estimator, as well as to be able to compute the CRLB, the derivative of the model Eq. (7.10) with respect to $\boldsymbol{\theta}$ is required. With the parameterization Eq. (7.13), the derivatives of the

model are given by

$$\frac{\partial A(\boldsymbol{q};\boldsymbol{\theta})}{\partial A_0} = \frac{\partial}{\partial A_0} A_0 \exp\left(-b\sum_{i,j=1}^{3} g_i g_j D_{ij}^{\text{app}} + \frac{b^2}{6}\left(\sum_{i=1}^{3}\frac{D_{ii}^{\text{app}}}{3}\right)^2 \sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l W_{ijkl}^{\text{app}}\right)$$
$$= \exp\left(-b\sum_{i,j=1}^{3} g_i g_j D_{ij}^{\text{app}} + \frac{b^2}{6}\left(\sum_{i=1}^{3}\frac{D_{ii}^{\text{app}}}{3}\right)^2 \sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l W_{ijkl}^{\text{app}}\right)$$
$$= \frac{A(\boldsymbol{q};\boldsymbol{\theta})}{A_0} \tag{7.15}$$

$$\frac{\partial A(\boldsymbol{q};\boldsymbol{\theta})}{\partial \boldsymbol{d}} = A(\boldsymbol{q};\boldsymbol{\theta})\left(-\sum_{i,j=1}^{3} q_i q_j \frac{\partial D_{ij}^{\text{app}}}{\partial \boldsymbol{d}} + \frac{2}{18}\frac{\sum_{i=1}^{3} D_{ii}^{\text{app}}}{3}\sum_{i,j,k,l=1}^{3} q_i q_j q_k q_l W_{ijkl}^{\text{app}} \sum_{i=1}^{3}\frac{\partial D_{ii}^{\text{app}}}{\partial \boldsymbol{d}}\right) \tag{7.16}$$

$$\frac{\partial A(\boldsymbol{q};\boldsymbol{\theta})}{\partial \boldsymbol{w}} = A(\boldsymbol{q};\boldsymbol{\theta})\frac{1}{6}\left(\sum_{i=1}^{3}\frac{D_{ii}^{\text{app}}}{3}\right)^2 \sum_{i,j,k,l=1}^{3} q_i q_j q_k q_l \frac{\partial W_{ijkl}^{\text{app}}}{\partial \boldsymbol{w}}, \tag{7.17}$$

$$\frac{\partial A(\boldsymbol{q};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left[\frac{\partial A(\boldsymbol{q};\boldsymbol{\theta})}{\partial A_0} \quad \frac{\partial A(\boldsymbol{q};\boldsymbol{\theta})}{\partial \boldsymbol{d}^T} \quad \frac{\partial A(\boldsymbol{q};\boldsymbol{\theta})}{\partial \boldsymbol{w}^T}\right]^T \quad , \tag{7.18}$$

where $\frac{\partial D_{ij}^{\text{app}}}{\partial d_k}$ and $\frac{\partial W_{ijkl}^{\text{app}}}{\partial w_m}$ is either 1 or 0, depending on whether the element from $\boldsymbol{d}$ or $\boldsymbol{w}$ fills that specific element of $\boldsymbol{D}$ or $\boldsymbol{W}$ or not.

Weighted least squares The ML estimator computes the optimum of a non linear function. Most efficient non linear optimization routines are local optimization routines. These routines require an initialization which is as good as possible for two reasons. First, to avoid convergence to potential erroneous local minima. Second, to converge with the fewest number of iterations. Therefore, we developed a good initialization for the ML estimator of the diffusion kurtosis model. This initialization is obtained with the (weighted) least squares estimator. First, to remove the exponential function, the logarithm of both sides of the DKI model Eq. (7.10) is taken:

$$\log A(\boldsymbol{q};\boldsymbol{\theta}) = \log(A_0) - b\sum_{i,j=1}^{3} g_i g_j D_{ij}^{\text{app}} + \frac{b^2}{6}\sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l M_{ijkl}^{\text{app}} \quad , \tag{7.19}$$

which can be written as:

$$\tilde{\boldsymbol{A}} = \boldsymbol{X}\tilde{\boldsymbol{\theta}}, \tag{7.20}$$

with

$$\tilde{\boldsymbol{\theta}} = \begin{bmatrix} \log(A_0) & \boldsymbol{d}^T & \tilde{\boldsymbol{w}}^T \end{bmatrix}^T \tag{7.21}$$

$$\tilde{\boldsymbol{w}} = \boldsymbol{w} \left(\sum_{i=1}^{3} \frac{D_{ii}^{\text{app}}}{3} \right)^2 \tag{7.22}$$

$$\tilde{\boldsymbol{A}} = \log(\boldsymbol{A}) \tag{7.23}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & (-b g_i g_j)_{1 \le i \le j \le 3} & \frac{b^2}{24} (g_i g_j g_k g_l)_{1 \le i \le j \le k \le l \le 3} \end{bmatrix} \tag{7.24}$$

where $\boldsymbol{A} = (N \times 1)$ is a vector containing the model predicted values for all DWI in a specific voxel, $\boldsymbol{X}$ $(N \times 22)$ is the regression matrix, where the index values i, j, k, l correspond to those of $\boldsymbol{\theta}$. To obtain the WLS estimate of the parameters $\tilde{\boldsymbol{\theta}}$, a sufficient number DWI, with sufficiently different gradient directions $\boldsymbol{g}$ and b-values vectors need to be recorded. When $\tilde{\boldsymbol{A}}$ contains the observed intensity of all DWI in a specific voxel, the parameters $\tilde{\boldsymbol{\theta}}$ of the diffusion process in that voxel can be estimated with :

$$\boldsymbol{W} = \text{diag}(\boldsymbol{A}^2) \tag{7.25}$$

$$\hat{\boldsymbol{\theta}}_{\text{LS}} = \left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{W} \tilde{\boldsymbol{A}} \tag{7.26}$$

where $\boldsymbol{W}$ is the inverse of the covariance matrix of the (independent) DWI measurements. This estimator assumes the noise level to be equal in each DWI. However, due to the log function, the variance of $\tilde{\boldsymbol{A}}$ is (approximately) $1/\boldsymbol{A}^2$ times the variance of $\boldsymbol{A}$.

As this estimator is substantially faster to compute, and has been observed to provide reasonably good results, it might even be used without the subsequent ML estimator.

7.2.2 CRLB of the kurtosis

The goal in this chapter is to optimize the experimental design of a diffusion weighting acquisition scheme such that diffusion kurtosis parameters can be estimated as precisely as possible. For this purpose we will employ the Cramér-Rao lower bound (CRLB) framework. As explained in Section 2.7, the CRLB provides a lower bound on the variance of any unbiased estimator $\hat{\boldsymbol{\theta}}$ of the parameters $\boldsymbol{\theta}$ of a statistical model of measurements. Experiments showed that the number of observations available in typical DKI measurements is sufficient for the asymptotic properties of the ML estimator, i.e. that it reaches the CRLB, to be valid.

The parameters of the DKI model are not directly relevant for analysis, but derived parameters, of which several will be explained in the next section, are. In order to optimize the CRLB of a derived kurtosis parameter, the CRLB

of this derived parameter need to computed. Let $m(\boldsymbol{\theta})$ be a derived kurtosis parameter, given as function of the DKI model parameters $\boldsymbol{\theta}$. Then, the CRLB of $m(\boldsymbol{\theta})$ is given by

$$\text{cov}(m(\hat{\boldsymbol{\theta}})) \geq I_m(\boldsymbol{\theta})^{-1}, \tag{7.27}$$

where I_m is given by

$$I_m(\boldsymbol{\theta})^{-1} = \frac{\partial m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \boldsymbol{I}(\boldsymbol{\theta})^{-1} \frac{\partial m(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad , \tag{7.28}$$

with $I(\boldsymbol{\theta})$ given by Eq. (2.33). The gradient settings of the DW-MR images can be optimized by minimizing the CRLB of a well chosen parameter $m(\boldsymbol{\theta})$.

7.2.3 Kurtosis parameters

In this section, the kurtosis parameters that will be considered in the experimental design and analysis are described.

Mean kurtosis

The mean kurtosis is given by

$$MK = \frac{1}{4\pi} \iint_{\boldsymbol{n}\in\mathbb{S}^2} K(\boldsymbol{n}) d\boldsymbol{n}, \tag{7.29}$$

where $K(\boldsymbol{n})$ is the parameterized excess kurtosis, given by

$$K(\boldsymbol{n}) = \frac{\left(\sum_{i=1}^{3} \frac{D_{ii}}{3}\right)^2 \sum_{i,j,k,l=1}^{3} n_i n_j n_k n_l W_{i,j,k,l}}{(\sum_{i,j=1}^{3} n_i n_j D_{i,j})^2}, \tag{7.30}$$

integrated over the unit sphere $\mathbb{S}^2$.The derivation of Eq. (7.30) is given in Appendix 7.A. Note that the definition of the mean kurtosis in Eq. (7.29) differs from a previous definition in Eq. {2} in [78], where the MK is computed by averaging the kurtosis in the sampled gradient directions. This requires the same gradient directions to be sampled at multiple b-values. We prefer the definition in terms of the integral since this allows free selection of the gradient directions and b-values and allows accurate mean kurtosis values, even when the gradient directions are not uniformly distributed on $\mathbb{S}^2$.

Radial kurtosis

The radial kurtosis is the mean of the kurtosis in the directions orthogonal to the direction of main diffusion:

$$K_\perp = \int_{\phi=0}^{2\pi} K\left(\boldsymbol{v}_2 \cos\phi + \boldsymbol{v}_3 \sin\phi\right) d\phi, \tag{7.31}$$

where $\boldsymbol{v}_i$ is the i^{th} eigenvector of the diffusion tensor $\boldsymbol{D}$, sorted by decreasing eigenvalue. This definition differs slightly from the radial kurtosis defined in [78], where the kurtosis along the second and third eigenvectors is averaged. The radial kurtosis is an interesting parameter, since the diffusion is restricted mainly in the radial direction. Therefore, it can be expected that the kurtosis, which is non zero due to the restricted diffusion, is most pronounced in the radial direction.

Kurtosis Anisotropy

In Eq. {8} in [78], the kurtosis anisotropy was defined as:

$$FA_K = \sqrt{\frac{2}{3}\frac{(K_1 - \bar{K})^2 + (K_2 - \bar{K})^2 + (K_3 - \bar{K})^2}{K_1^2 + K_2^2 + K_3^2}}, \tag{7.32}$$

with $K_i \equiv K(\boldsymbol{v}_i)$ and $\bar{K} = (K_1 + K_2 + K_3)/3$ the mean of the kurtosis in the diffusion tensor eigenvector directions. This definition is in direct analogy to the fractional anisotropy in DTI. The original motivation for the diffusion fractional anisotropy (FA) was that it is a coordinate system invariant, dimensionless characterization of the differences between diffusion in the different directions. However, in contrast to the diffusion eigenvalues, the kurtosis itself is dimensionless and thus does not need to be normalized to obtain a dimensionless value. In our opinion, a kurtosis anisotropy parameter should not scale with the inverse mean kurtosis, which might be zero, but should only be based on the variability in the kurtosis. Also, note that Eq. (7.32) only uses the kurtosis in the 3 directions specified by the diffusion eigenvectors. Since the kurtosis is specified by a higher order tensor, this might not capture all kurtosis variability.

We propose a different kurtosis anisotropy parameter, which, in our opinion, is more in line with the important characteristics of the FA. This new kurtosis anisotropy (KA) parameter is given by the standard deviation of the kurtosis:

$$KA = \sqrt{\frac{1}{4\pi}\iint_{\boldsymbol{n}\in\mathbb{S}^2} \left(K(\boldsymbol{n}) - MK\right)^2 d\boldsymbol{n}}. \tag{7.33}$$

Efficient evaluation of the integrals of the kurtosis measures

The integral in Eq. (7.29) and Eq. (7.33) over $\mathbb{S}^2$ is numerically approximated by sampling the integrand on a weighted set of n, approximately uniformly distributed, sample points $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n$ in $\mathbb{S}^2$. The weights $\boldsymbol{w} = [w_1, \cdots, w_n]$ of the sample points are chosen to maximize the accuracy of the approximation of the integrals. For this, first note that any function on $\mathbb{S}^2$ can be expanded in the spherical harmonics basis. Since the integral of all non zeroth order spherical harmonics is zero, integrating over $\mathbb{S}^2$ determines the magnitude of the zeroth

order spherical harmonic of the integrand, multiplied by the area of $\mathbb{S}^2$. As the zeroth order spherical harmonic is constant, accurate evaluation of the integral requires that $\sum_i w_i = 4\pi$. Smooth functions, such as the integrands considered in Eq. (7.29) and Eq. (7.33), generally have a decaying spectrum; i.e. the energy of the functions is mainly concentrated in low order spherical harmonics. To avoid contamination of the evaluated integral by the energy in these low order spherical harmonics, the weights $\boldsymbol{w}$ are orthogonalized to a finite number of spherical harmonics. To minimize the influence of energy of the integrands in arbitrary higher order harmonics, the 2-norm of $\boldsymbol{w}$ is minimized. It can be proven that the weight vector $\boldsymbol{w}$ that has these properties is the solution of a least squares problem with only 1 non-confound:

$$\boldsymbol{w} = [4\pi \ , \ 0 \ , \dots \ , 0] \left(\boldsymbol{S}^T \boldsymbol{S}\right)^{-1} \boldsymbol{S}^T, \tag{7.34}$$

where the components of $\boldsymbol{S}$ are given by:

$$S_{i,j} = Y_{l_j}^{m_j}(\boldsymbol{p}_i) \tag{7.35}$$

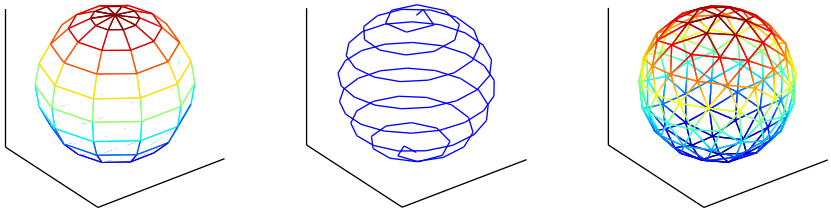
where $Y_{l_j}^{m_j}$ is the m_j^{th} real valued spherical harmonic of order l_j, and $l_1 = 0$, $m_1 = 0$. Usually, all $(\mathfrak{l}+1)^2 < n$ unique combinations of $|m_j| \leq l_j \leq \mathfrak{l}$ have to be included, where $\mathfrak{l}$ is the maximum order of spherical harmonics to which $\boldsymbol{w}$ is orthogonal. However, when both the set of points and the integrand are symmetric around 0, the odd l_j and a symmetric half of the points $\boldsymbol{p}_i$ do not need to be included in the computations. The integrands in Eq. (7.29) and Eq. (7.33) are symmetric around 0, so by selecting a symmetric set of points, this property can be used.

Distributing point uniformly on a sphere

As described in the previous subsection, the integration over the spere requires a set of points that are approximately isotropically distributed over this sphere. Since the surface of the sphere is a curved space, a set of points cannot be distributed evenly, except for the sets of points defined by the platonic solids, containing 4, 6 8, 12 or 20 points. Obviously non uniform covers of the sphere can be created:

1. First it is possible to space the points on a regular regular in spherical coordinates. This gives a distribution like the often used intersections of the circles of constant longitude or latitude of the earth. This distribution gives a substantially increased density at the poles.

2. A second possibility, which already provides much more equal spacing, is by threading a spiral around the sphere and sampling at regular steps along this spiral.



(a) (b) (c)

Figure 7.1: *Meshes of thee different covers of the sphere with 100 points. (a) regular grid in spherical coordinates, (b) regularized spiral, (c) points optimized with electrostatic potential.*

3. The third option, explained in more detail in this section, is by optimizing a set of points, by simulating repulsion between the points.

See Fig. 7.1 for an example of these three distributions, where sets with 100 points were selected.

Distribution 1: For the regular sampling in the spherical coordinates, the distribution the points P is created by sampling θ from $-\pi$ to π with $n_\theta + 1$ equally spaced samples and ϕ from $-\pi/2$ to $\pi/2$ with $n_\phi + 1$ samples. When double points are removed, $n = n_\theta(n_\phi - 1) + 2$ points remain.

Distribution 2: The spiral sampling of the sphere was defined in [79]. This is a generalized spiral set of points that tries to approximate optimal isotropic sets of points.

Distribution 3: The optimized sets of points. In general the points can be optimized by minimizing

$$\boldsymbol{P} = \arg\min \sum_{1 \leq i < j \leq n} |\boldsymbol{p}_i - \boldsymbol{p}_j|^\alpha \tag{7.36}$$

for some power α of the Euclidean distance between the points $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$. Since the optimization involves many degrees of freedom, $2n - 3$, the optimization consumes quite some computation time and is best evaluated offline. Since the main specifying parameter is the number of points n, only a relative low number of optimized sets need to be stored.

The optimization itself is not trivial due to the many degrees of freedom and the presence of a large number of local minima. It is practically impossible to guarantee convergence to the global minimum when more than, say, 30 points are optimized. However, by optimizing with multiple initializations, a very close to optimal set of points can be found. The quality of the set of optimized

points could potentially depend on the kind of application the set of points is optimized for. However, it was observed that the quality for integration does not strongly depend on α, as long as it is negative. Therefore, α can be chosen such that the function can be efficiently evaluated. The most easy and fast to evaluate powers are integer powers close to 0, such as -1 or -2. Note that for $\alpha < -2$ the function value is more strongly influenced by the nearest neighbors, in effect maximizing the distance to the nearest neighbors. General non linear optimization routines strongly benefit from the availability of the gradient of the function to optimize. Fortunately, computing the gradient of $|\boldsymbol{p}_i - \boldsymbol{p}_j|^\alpha$ is almost trivial. In our implementation, which is in a MATLAB routine called `sphere_best.m`, a limited memory BGFS optimization method was employed for the actual non linear optimization.

7.2.4 Optimization of the gradient settings

Let the settings of all diffusion weighting gradients during a DKI experiment, in which N DW-MR images are acquired, be defined by $\boldsymbol{Q} = [\boldsymbol{q}_1, \dots, \boldsymbol{q}_N]^T$. Here each $\boldsymbol{q}_i$ specifies the gradient settings, i.e. the b-value and the direction of the diffusion weighting gradient, of a DW-MR image. Then, $\boldsymbol{Q}$ can be optimized by minimizing the CRLB of the model parameters, $\boldsymbol{I}(\boldsymbol{\theta})^{-1}$. However, optimization methods need a scalar function to optimize. Therefore, a scalar objective function of the CRLB of the model parameters is required. This scalar objective function should evaluate the overall quality of the gradient settings $\boldsymbol{Q}$. Our objective function is the CRLB of a kurtosis parameter, $I_m(\boldsymbol{\theta})^{-1}$, given in Eq. (7.28). This function depends on the actual tissue properties $\boldsymbol{\theta}_0$. These properties $\boldsymbol{\theta}_0$ are generally different in each voxel. Since the brain images contain many voxels and different tissues in which one might be interested, the acquisition scheme should be optimal for a distribution of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$. The optimal $\boldsymbol{Q}$ can then be found by minimizing the objective function, weighted with the prior distribution $p(\boldsymbol{\theta})$:

$$\hat{\boldsymbol{Q}} = \arg\min_{\boldsymbol{Q}} \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}) I_m(\boldsymbol{\theta})^{-1} d\boldsymbol{\theta}. \tag{7.37}$$

In practice, the prior distribution $p(\boldsymbol{\theta})$ can be approximated with experimental data. However, it is difficult to evaluate the 22 dimensional integral in Eq. (7.37). Therefore, it often is much more convenient to approximate the integral by drawing M samples from the prior distribution $p(\boldsymbol{\theta})$ and evaluating the objective function on this set only. These M samples $\boldsymbol{\theta}_i$ from $p(\boldsymbol{\theta})$, which should be fixed during the optimization, can be collected in a set $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_M\}$. Then, the integral in Eq. (7.37) can be approximated by

$$H(\boldsymbol{Q}; \boldsymbol{\Theta}) = \frac{1}{M} \sum_{i=1}^{M} I_m(\boldsymbol{\theta}_i)^{-1}, \tag{7.38}$$

which is the mean CRLB of the kurtosis parameter of the elements of $\boldsymbol{\Theta}$. The optimal gradient settings are then given by

$$\hat{\boldsymbol{Q}} = \arg\min_{\boldsymbol{Q}} H(\boldsymbol{Q}; \boldsymbol{\Theta}). \tag{7.39}$$

7.2.5 Efficient implementation of the optimization

The optimal set $\boldsymbol{Q}$ is in principle given by Eq. (7.39). However, due to the large number of diffusion weighted images in a typical DKI experiment, it is not trivial to find this optimum. The actual function to optimize is a sum of scalar functions, each of which depends on the CRLB of a parameter vector $\boldsymbol{\theta}$, which is a 22×22 matrix. Most common optimization techniques for large problems use the analytic or numerically computed derivatives of the function to be optimized. However, the derivative of the inverse Fisher matrix $\boldsymbol{I}$ with respect to $\boldsymbol{Q}$ is difficult and computationally expensive to compute. There are general optimization routines which do not use the gradient of the function, for example `fminsearch` in MATLAB (The MathWorks, Inc.), but they typically require a huge number of function evaluations and it was observed that the final gradient set found by `fminsearch` was not close to optimal. This might be due to local minima and/or almost flat parts in the objective function. To be able to overcome local minima, simulated annealing was chosen as alternative optimization method [**?**]. The simulated annealing method iteratively updates the diffusion weighting gradients, one at a time. When a gradient is modified, the Fisher information matrix $\boldsymbol{I}$ needs to be updated to evaluate the change in objective function. From Eq. (2.33) it follows that updating the magnitude and direction of one diffusion weighting gradient, which influences the magnitude in one MR image, causes a rank two update of $\boldsymbol{I}$. Therefore, the Woodbury identity [**?**],

$$(A + CBC^T)^{-1} = A^{-1} - A^{-1}C\left(B^{-1} + C^T A^{-1} C\right)^{-1} C^T A^{-1}, \tag{7.40}$$

can be used to efficiently update the inverse Fisher matrix.

7.3 Experiments

As described in Section 7.2.4, the gradient settings are optimized by minimizing the CRLB of a kurtosis parameter, evaluated on a set $\boldsymbol{\Theta}$ obtained from a prior distribution of DKI model parameters. In practice, samples from this distribution are obtained from prior DKI measurements. In this chapter, the prior DKI measurements were obtained from a human and small animal DKI experiments of which the details are described in Section 7.3.1. In Section 7.3.2, several aspects that are important for the selection of $\boldsymbol{\Theta}$ are discussed. Then, in section 7.3.3, the sets $\boldsymbol{\Theta}$ that were used for the optimizations are specified. Finally, in section 7.3.4, traditional gradient settings are reviewed and the settings of the optimized gradients are described.

7.3.1 Acquisition of DKI data

Human DKI data was used for the construction of $\mathbf{\Theta}$. This data was acquired after approval of the institutional review board and after informed consent was obtained from the healthy volunteer. The dual spin echo diffusion-weighted 2D EPI images were acquired with a Siemens 3.0-T MRI scanner. The volume was recorded with 45 slices with an acquisition matrix of 82 × 82. The voxel dimensions were 2.7 mm isotropic and the echo time was TE = 96 ms. A 8 channel head coil was used and the GRAPPA acceleration factor was 2, with 24 reference lines. The bandwidth was 1356 Hz/pixel and the TR = 5.900 sec. The maximum b-value of the set of recorded DW-MR images was 2800 s/mm^2. The SNR of the grey matter in the MR images with a b-value of zero was 12. To evaluate the performance and robustness of the optimized gradient settings, a second volunteer was scanned with the optimized settings on a different Siemens 3.0-T MRI scanner. This dataset was recorded with 55 slices with an acquisition matrix of 128 × 128. The voxel dimensions were 2.5 mm isotropic and the echo time was TE = 104 ms. A 30 channel head coil was used and the GRAPPA acceleration factor was 2, with 24 reference lines. The bandwidth was 1955 Hz/pixel and the TR = 7.700 sec. The maximum b-value of the set of recorded DW-MR images was 2800 s/mm^2. Furthermore, DKI data of a rat was acquired with a Bruker 7T small animal scanner. This dataset was acquired with 50 slices with an acquisition matrix of 96 × 64, reconstructed to an image size of 128 × 64. The slice thickness was 0.37 mm, excluding the gap of 0.10 mm between slices. The in plane resolution was 0.37 mm and the echo time was TE = 24 ms. The bandwidth was 8333 Hz/pixel and the TR = 11.000 sec. The images were recorded with 2 shot EPI and mono polar diffusion weighting gradients with a maximum b-value of 2800 s/mm^2, obtained with the diffusion times $\delta = 5ms$ and $\Delta = 12ms$. From these datasets, the model parameters were estimated with a Maximum Likelihood estimator [14].

7.3.2 Prerequisites for the selection of prior DKI model parameters

There are several aspects that influence the selection of $\mathbf{\Theta}$:

- For elements of $\mathbf{\Theta}$ with a large positive kurtosis, the magnitude of the DW-MR images, predicted with the DKI model Eq. (7.10), will grow strongly for large q. This is caused by the 4^{th} power of q inside the exponent function and indirectly, it is a result of the finite region in which the series expansion used for the DKI model is accurate. The high SNR of these anomalously high predicted DW-magnitudes reduce the CRLB of the kurtosis parameters, which might cause the optimization procedure to increase some gradients to unrealistically high magnitudes. This can be avoided by limiting the maximum q value that is allowed in

the optimization or by selecting elements for $\boldsymbol{\Theta}$ without large positive kurtosis.

- The DKI parameters will depend on the tissue type under study. Therefore, representative parameter vectors of the different tissues should be included in $\boldsymbol{\Theta}$. When the set $\boldsymbol{\Theta}$ is too small or does not contain elements from all relevant tissue types, the optimized gradient settings might be good for the kurtosis parameters in the test set, but not for all the various brain tissue types.
- The computation time required for the optimization depends almost linearly on the number of elements in $\boldsymbol{\Theta}$. Therefore, to limit the computation time required by the optimization, $\boldsymbol{\Theta}$ should not contain excessively many elements. In our experiments, optimizations are performed with several hundreds of test tensors in $\boldsymbol{\Theta}$.
- Some kurtosis parameters, such as the radial kurtosis, depend on the intrinsic diffusion tensor coordinate system. When the optimization is performed w.r.t. these kurtosis parameters, the coordinate system should be well defined for all elements of $\boldsymbol{\Theta}$. This can be established by selecting $\boldsymbol{\Theta}$ from sufficiently anisotropic tissues, such as the white matter structures.

7.3.3 Selection of DKI model parameter sets

This section describes the sets $\boldsymbol{\Theta}$. These sets $\boldsymbol{\Theta}$ contain the samples from the prior distribution $p(\boldsymbol{\theta})$ that were selected for the optimization experiments. Each element of a $\boldsymbol{\Theta}$ contains the parameters of the DKI model, from which the kurtosis parameters can be evaluated. In order to investigate the sensitivity of the optimization of the kurtosis parameters to different $\boldsymbol{\Theta}$, the optimization is performed for the following three sets:

$\boldsymbol{\Theta}_1$ To avoid unrealistically large b-values due to large positive kurtosis, the set $\boldsymbol{\Theta}_1$ was constructed to have **zero kurtosis** and a range of realistic diffusion tensor eigenvalues. The diffusion eigenvalues were typical for the grey matter, white matter, and the cerebrospinal fluid (CSF) present in the human DKI dataset. To avoid indeterminacy of diffusion tensor eigenvectors, the eigenvalues were chosen sufficiently different. The diffusion tensor eigenvalues are given in Table 7.2 and were manually selected from the DKI dataset (cfr. Section 7.3.1). Furthermore, to make the gradient settings to be optimized (approximately) rotationally invariant, 60 diffusion tensors were generated from each set of eigenvalues by rotating the first eigenvector towards the 20 corners of a dodecahedron and subsequently rotating the second eigenvector in steps of 120°.

$\boldsymbol{\Theta}_2$ The second set was obtained by randomly selecting 400 DK-tensors from the **white and grey matter** of the DKI dataset. The probability to

be included in the set was equal for each grey and white matter voxel. Since the estimated parameter vectors are noise corrupted, the diffusion eigenvalues might occasionally be unrealistically low or the estimated kurtosis might be large in some directions. Therefore, the lowest diffusion eigenvalues were adjusted to be at least $1/b_m$, with $b_m = 3000$ s/mm^2 and for all directions $\boldsymbol{g}$ in which the kurtosis was positive, the kurtosis was decreased as long as $q_m^2 t < b_m$, with $q_m = \arg\min_q A(q\boldsymbol{g})$, with A from Eq. (7.10).

$\boldsymbol{\Theta}_3$ The third set was obtained by randomly selecting 400 DK-tensors from **white matter only**, FA>0.4, of the DKI dataset. The probability of a voxel to be included in the selection was proportional to the FA value of the voxel. The elements of $\boldsymbol{\Theta}_3$ were adjusted with the same procedure as the elements of $\boldsymbol{\Theta}_2$.

Table 7.2: *Diffusion eigenvalues for $\boldsymbol{\Theta}_1$ in mm^2/s and FA value*

$\lambda_1 \times 10^3$	$\lambda_2 \times 10^3$	$\lambda_3 \times 10^3$	FA
0.955	1.076	1.287	0.1507
0.940	1.103	1.309	0.1639
0.525	0.687	2.813	0.7514
0.631	1.250	1.839	0.4528
0.522	0.680	2.570	0.7286
0.525	0.782	2.317	0.6711
4.057	4.349	4.877	0.0936
3.217	3.441	3.944	0.1050

7.3.4 Optimized and traditional gradient settings

The optimized gradient settings were compared with a 'traditional' set T of diffusion gradient settings for DKI [71]:

T specifies DW-MR images with diffusion weighting gradients in 30 directions, with 5 different b-values, 500 to 2500 s/mm^2 in steps of 500 s/mm^2 and 10 images with a b-value of zero. So in total T specifies 160 DW-MR images.

For fair comparisons, the optimized sets used the same number of DW-MR images as the traditional set T. The optimized sets of gradient settings were:

$O_{i,m}^{b_j}$ These DKI gradient settings were optimized with $\boldsymbol{\Theta}_i$, with a maximum b-value b_j allowed in the optimization, with $b_1 = 2800$ s/mm^2 and $b_2 = 2500$ s/mm^2. The optimization was performed with respect to the CRLB of

the mean kurtosis of the elements of $\boldsymbol{\Theta}_i$. The maximum allowed b-value was limited to avoid the selection of excessively large b-values due to the breakdown of the DKI model for very high b-values.

$O_{i,r}^{b_j}$ The same as $O_{i,m}^{b_j}$, but now optimized with respect to the CRLB of the radial kurtosis obtained from $\boldsymbol{\Theta}_i$.

7.4 Results and discussion

In this section, the results of the experiments are discussed. First, in Section 7.4.1, the optimized and traditional gradient settings are compared on kurtosis parameters of the sets $\boldsymbol{\Theta}$. Next in Section 7.4.2, a good optimized set of gradient settings is reviewed. Finally, in Section 7.4.3, recorded DKI data is used to compare the performance of the optimized gradient settings with the traditional gradient settings.

7.4.1 Results of the optimization

In this section, the performance of the different gradient settings is compared.

Table 7.3 shows the relative $H(\boldsymbol{Q};\boldsymbol{\Theta}_i)$, Eq. (7.38). That is, Table 7.3 shows the normalized mean CRLB of the mean (Table 7.4a) or radial (Table 7.4b) kurtosis from the three sets $\boldsymbol{\Theta}_i$, for all gradient sets $\boldsymbol{Q}$. The values are normalized by dividing with the lowest $H(\boldsymbol{Q};\boldsymbol{\Theta}_i)$ in each column. This table clearly shows that the gradient settings influence the precision with which the kurtosis parameters can be estimated, as the normalized mean CRLB of kurtosis parameters of the elements of $\boldsymbol{\Theta}_i$ is different for each gradient set. Also, Table 7.3 shows that only the gradient settings optimized for $\boldsymbol{\Theta}_1$, i.e. $O_{1,r}^{b_1}$ and $O_{1,m}^{b_1}$, have a low relative mean CRLB on kurtosis parameters computed from $\boldsymbol{\Theta}_1$. The relative mean CRLB of the other gradient settings, $O_2^{b_j}$ and $O_3^{b_j}$, on the estimation of kurtosis parameters of $\boldsymbol{\Theta}_1$ is much larger.

Further inspection showed that these other gradient settings had a substantially higher CRLB on the elements of $\boldsymbol{\Theta}_1$ that modeled CSF. This is caused by the very high diffusivity of CSF, which is not present in the grey or white matter from which $\boldsymbol{\Theta}_2$ and $\boldsymbol{\Theta}_3$ were selected. Since one is usually not interested in the CSF, this is not a problem for the use of the other optimized gradient settings, but clearly shows the importance of the selection of the elements of $\boldsymbol{\Theta}$.

Furthermore, Table 7.3 shows that the traditional gradient settings T has a higher mean CRLB, especially for $\boldsymbol{\Theta}_3$, which contains parameter vectors of white matter structures. The mean CRLB of the radial kurtosis of T is 2.5 times larger than the mean CRLB of the radial kurtosis of the best gradient settings.

Table 7.3 further shows that the gradient settings $O_{2,m}^{b_1}$, optimized for the mean kurtosis based on $\boldsymbol{\Theta}_2$, generalizes well. That is, the mean CRLB of the

mean and radial kurtosis of $\mathbf{\Theta}_2$ and $\mathbf{\Theta}_3$ is close to minimal, when the gradient settings are specified by $O^{b_1}_{2,m}$. Since $\mathbf{\Theta}_2$ was randomly selected from the grey and white matter, it contains mostly grey, but also white matter voxels, which might explain the relatively low mean CRLB of the kurtosis parameters on the white matter only set $\mathbf{\Theta}_3$.

The sets of gradient settings $O^{b_2}_i$ were limited to a maximum b-value of 2500 s/mm^2. When the performance of these sets is compared to $O^{b_1}_i$, it is clear that the mean CRLB for the set $\mathbf{\Theta}_i$ which they are optimized on is only slightly increased. In particular, the mean CRLB of the mean kurtosis of $\mathbf{\Theta}_2$ is only increased by 6.1% for $O^{b_2}_{2,m}$, compared to $O^{b_1}_{2,m}$. However, the mean CRLB of different kurtosis parameters or model parameters is increased substantially by this lower b-value of 2500 s/mm^2. This can, for example, be seen by comparing the relative mean CRLB of the radial kurtosis of $\mathbf{\Theta}_3$ of the gradient settings $O^{b_2}_{2,m}$ with that of $O^{b_1}_{2,m}$, i.e. 1.758/1.379, which is larger than the 6.1% increase in mean CRLB of the mean kurtosis of $\mathbf{\Theta}_2$ for these gradient settings.

Summarized, the tables 7.3 and 7.4b show that:

- The mean CRLB can be substantially decreased by optimizing the set of gradients.
- For the optimization, it is important that the selection of samples from a prior distribution of tensors matches the diffusion and kurtosis properties found in the relevant tissues under study.
- Gradient sets with good performance for both grey and white matter structures and for both the mean and radial kurtosis can be found, such as our set $O^{b_1}_{2,m}$

7.4.2 Optimized set of DW-gradients

From the results presented in the previous section, it can be concluded that the optimized gradient set $O^{b_1}_{2,m}$ produced the best results overall. Therefore, Fig. 7.2 shows the optimized gradients $O^{b_1}_{2,m}$, which are optimized with respect to the mean kurtosis of $\mathbf{\Theta}_2$. Fig. 7.2a shows the sorted b-values of the optimized gradient set. The b-values automatically separate in, more or less, distinct levels. Fig. 7.2b shows the gradient directions and magnitudes of the individual directions, plotted on a sphere. The density of gradients is indicated by the grey level of the sphere. As can be seen in Fig. 7.2b, the gradient directions are approximately isotropically distributed, which is a result that is obtained automatically by the optimization.

As in the acquired DKI dataset, the maximum b-value allowed in the optimization was 2800 s/mm^2. Fig. 7.2a shows that a substantial number of gradients are located at this maximum b-value. This suggests that the precision might be improved by increasing the maximum allowed b-value even further.

Table 7.3: *Normalized value of the objective function H of the mean (a) kurtosis, radial (b) kurtosis, or kurtosis anisotropy (c). This objective function H is the average CRLB of the mean or radial kurtosis of the gradient sets for the different test sets $\mathbf{\Theta}_i$. The normalization was performed such that the lowest value in each column was 1.*

	$\mathbf{\Theta}_1$	$\mathbf{\Theta}_2$	$\mathbf{\Theta}_3$
T	151	1.406	1.911
$O^{b_1}_{1,m}$	1	2.299	4.737
$O^{b_1}_{1,r}$	1	1.455	1.880
$O^{b_1}_{2,m}$	496	1.000	1.186
$O^{b_1}_{2,r}$	116	1.261	1.437
$O^{b_1}_{3,m}$	203	1.367	1.000
$O^{b_1}_{3,r}$	453337	3.983	1.137
$O^{b_2}_{2,m}$	211	1.061	1.396
$O^{b_2}_{2,r}$	125	1.296	1.653
$O^{b_2}_{3,m}$	326	1.584	1.242
$O^{b_2}_{3,r}$	460150	2.623	1.357

(a) Mean kurtosis

	$\mathbf{\Theta}_1$	$\mathbf{\Theta}_2$	$\mathbf{\Theta}_3$
T	77	1.571	2.541
$O^{b_1}_{1,m}$	1	1.966	7.173
$O^{b_1}_{1,r}$	1	1.398	2.267
$O^{b_1}_{2,m}$	332	1.410	1.379
$O^{b_1}_{2,r}$	80	1.000	1.480
$O^{b_1}_{3,m}$	260	1.719	1.112
$O^{b_1}_{3,r}$	220969	3.696	1.000
$O^{b_2}_{2,m}$	112	1.393	1.758
$O^{b_2}_{2,r}$	73	1.042	1.842
$O^{b_2}_{3,m}$	410	1.720	1.517
$O^{b_2}_{3,r}$	226377	2.517	1.406

(b) Radial kurtosis

	$\mathbf{\Theta}_2$	$\mathbf{\Theta}_3$
T	1.253	1.717
$O^{b_1}_{1,m}$	1.744	3.716
$O^{b_1}_{1,r}$	1.319	1.569
$O^{b_1}_{2,m}$	1.000	1.177
$O^{b_1}_{2,r}$	1.110	1.316
$O^{b_1}_{3,m}$	1.616	1.000
$O^{b_1}_{3,r}$	4.515	1.113
$O^{b_2}_{2,m}$	1.048	1.345
$O^{b_2}_{2,r}$	1.130	1.444
$O^{b_2}_{3,m}$	1.803	1.210
$O^{b_2}_{3,r}$	2.957	1.282

(c) kurtosis anisotropy

However, as the model Eq. (7.10) is based on a series expansion, it is not suited to extrapolate the magnitude of DW-MR images acquired with higher b-values.

7.4.3 Comparison of the precision

This section compares the CRLB of kurtosis parameters of the optimized gradient set $O_{2,m}^{b_1}$ with the traditional gradient set T. Fig. 7.3a and Fig. 7.3b show the mean and radial kurtosis of the human DKI dataset, respectively. As is clearly visible, the radial kurtosis is substantially larger in the white matter structures, compared to the mean kurtosis. This indicates that the deviations from the Gaussian distribution are strongest in the radial direction. Fig. 7.4 shows the square root of the CRLB of the mean kurtosis, evaluated with the traditional and the optimized gradient sets, respectively. Thus, this figure displays a lower bound on the standard deviation of the mean kurtosis. In Fig. 7.4, it is clearly visible that the kurtosis parameters of the CSF cannot be precisely estimated by both gradient settings, as the square root of the CRLB of the mean kurtosis is high compared to the mean kurtosis of the CSF. The CRLB of the other tissues is lower, indicating more precise estimates. In Fig. 7.4, the differences between the CRLB of the mean kurtosis of the gradient sets are difficult to see. Therefore, the precision of a kurtosis parameter estimate obtained with the gradient sets T and $O_{2,m}^{b_1}$ was compared by evaluating the logarithm of the ratio of the CRLB of that kurtosis parameter,

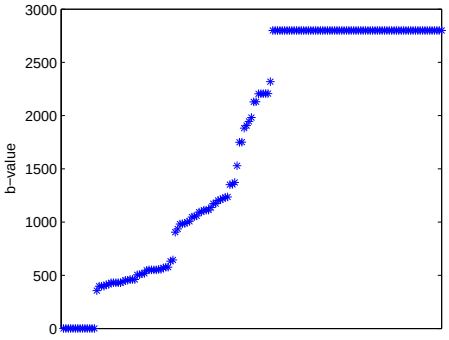
$$R(\boldsymbol{\theta}; T, O_{2,m}^{b_1}) = \ln\left(\frac{I_m^{-1}(\boldsymbol{\theta}, T)}{I_m^{-1}(\boldsymbol{\theta}, O_{2,m}^{b_1})}\right) \quad . \tag{7.41}$$

The value of R is 0 when both gradient sets have an equal CRLB of the kurtosis parameter in that voxel. Positive values indicate that the CRLB of the kurtosis parameter obtained with gradient settings T is larger than that obtained with the optimized gradient set $O_{2,m}^{b_1}$. Fig. 7.5 shows R of the mean kurtosis, radial kurtosis, and kurtosis anisotropy parameters. It is clearly visible that the gradient set $O_{2,m}^{b_1}$ improves the CRLB for all brain structures, except for the CSF. It was observed that the median reduction of the CRLB of the mean kurtosis in the grey matter was a factor 2.1. The factor is even larger than the value obtained in the simulation experiment with $\boldsymbol{\Theta}_2$, which consists of a selection of grey and white matter voxels. Fig. 7.6 shows $R(\boldsymbol{\theta}, T, O_{2,m}^{b_1})$ for the diffusion parameters. The median of the CRLB of the mean diffusion in the grey matter is 12% larger for the traditional gradient settings compared with $O_{2,m}^{b_1}$. The precision of the FA and direction of the first eigenvector is almost equal for these 2 sets of gradient settings. Fig. 7.5 also shows the difference in performance between the gradient sets optimized with respect to the mean and radial kurtosis. As could be expected, the CRLB of the mean kurtosis is higher for $O_{2,r}^{b_1}$ than for $O_{2,m}^{b_1}$. The median CRLB of the mean kurtosis in the grey matter increases by 29%, the radial kurtosis is approximately the same

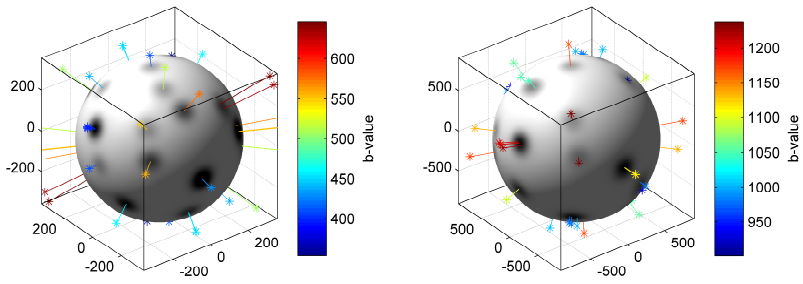
and the CRLB of the kurtosis anisotropy increases by 19%. This is a further indication that the gradient settings $O^{b_1}_{2,m}$ are better than $O^{b_1}_{2,r}$.
The last row of Fig. 7.5 shows $R(\boldsymbol{\theta}, T, O^{b_1}_{2,m})$ of the second dataset, recorded with the optimized gradient settings $O^{b_1}_{2,m}$ (accidently) rotated 90° in the slice direction. These figures are quite similar to the first row. This demonstrates that the optimized gradient settings are robust to differences between subjects and not very sensitive to changes in the acquisition parameters and orientation, as the acquisitions differ substantially. The median reduction of the CRLB of the mean kurtosis in the grey matter of this dataset was a factor of 1.9, which, considering the differences in acquisition, is close to the original improvement factor.
To further study the influence of changing recording settings, the DKI recording of a rat was made and the results are presented in Fig. 7.8. This figure shows that the gradient settings $O^{b_1}_{2,m}$, which are optimized for the human brain, improve the precision of the mean kurtosis, compared to the traditional gradient settings T. For the entire brain, the median $R(\boldsymbol{\theta}, T, O^{b_1}_{2,m}) = 0.72$, while the gradient settings optimized for this acquisition have a median $R(\boldsymbol{\theta}, T, O_{rat}) = 0.96$. So from these datasets we find that the gradient settings specific for the rat brain improve the CRLB of the median kurtosis by 27%, compared to the gradient settings found for the human brain. This relatively small difference, w.r.t. the substantially different MR system and subject, that the optimized gradient settings can be applied to slightly different acquisitions without substantial loss in precision of the diffusion and kurtosis parameters.
Finally, Fig. 7.7 shows the performance, as measured by $R(\boldsymbol{\theta}, T, *)$, of the optimized gradient settings $*$ as function of the number of elements in $\boldsymbol{\Theta}$. The value of R was computed for each $\boldsymbol{\theta}$ of the 20k voxels from the first dataset that were not used in any optimization. As is clearly indicated by this figure, this performance quickly levels off and is essentially constant above 350 elements. Hence, 400 elements are sufficient to optimize the gradient settings.

7.5 Conclusion

In summary, this chapter presents a novel method to optimize the diffusion weighting gradient settings. This method is based on the minimization of the Cramér Rao lower bound (CRLB) for estimating kurtosis parameters. The results show that the increase in precision that can be obtained, compared to a traditionally used set of gradients, is substantial. For the mean kurtosis estimated in grey matter voxels an improvement of the CRLB with a factor of 2.1 was observed. Alternatively, when the required precision is already achieved using standard gradient settings, optimizing the gradient settings allows one to achieve the same precision from a reduced number of DW MR images or from a set of DW-MR images with a reduced signal to noise ratio (allowing a higher resolution).



(a)



(b)

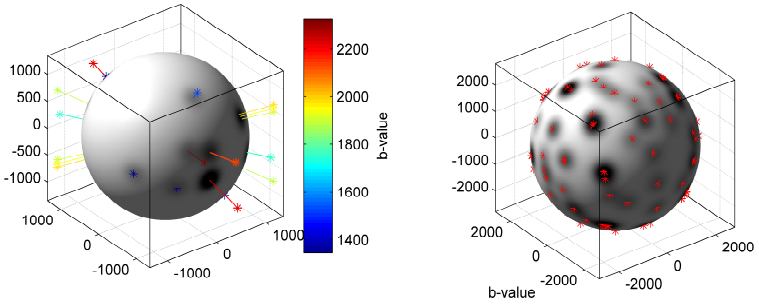
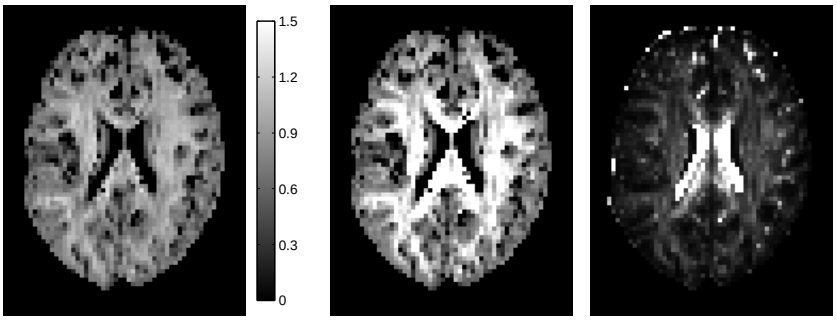
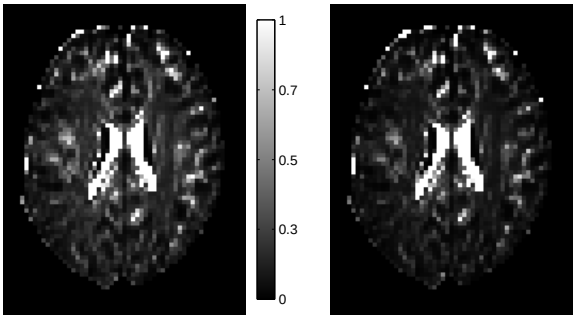


Figure 7.2: *Magnitude and direction of optimized gradients* $O_{2,m}^{b_1}$.



(a) Mean kurtosis (b) Radial kurtosis (c) Kurtosis anisotropy

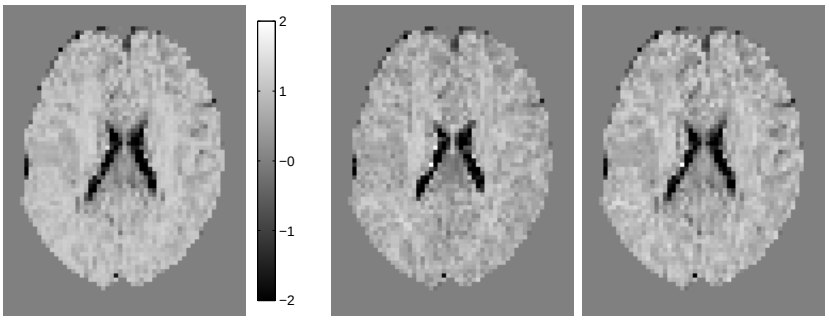
Figure 7.3: *Mean kurtosis, radial kurtosis and kurtosis anisotropy of a human volunteer.*



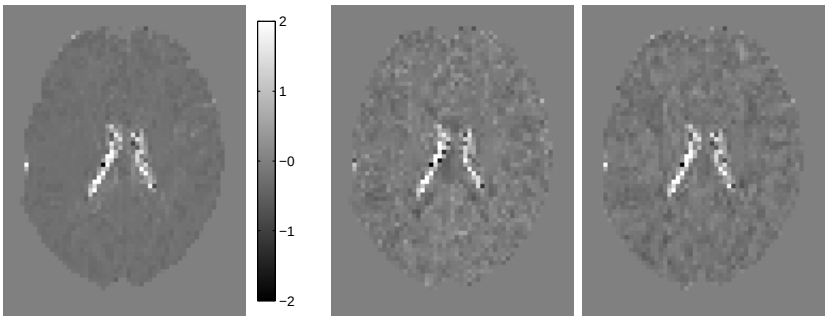
(a) $\sqrt{CRLB(T)}$ (b) $\sqrt{CRLB(O_{2,m}^{b_1})}$

Figure 7.4: $\sqrt{CRLB}$ *of the mean kurtosis, when estimated with* T *or* $O_{2,m}^{b_1}$.

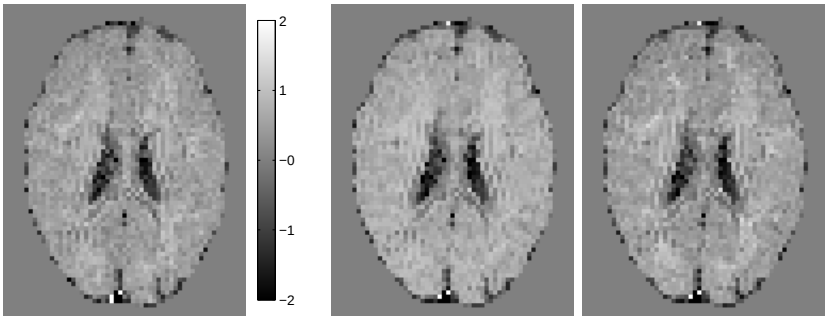
The gradient settings are optimized using a prior distribution of DKI model parameters. This prior distribution can be obtained from DW-MR images. In this chapter it was shown that the prior distribution substantially influences the precision of the kurtosis parameters. However, it was also shown that the performance of the optimized gradient settings is not substantially reduced when a different subject is scanned or other parameters of the acquisition are (slightly) changed.



(a) Mean kurtosis (b) Radial kurtosis (c) Kurtosis anisotropy

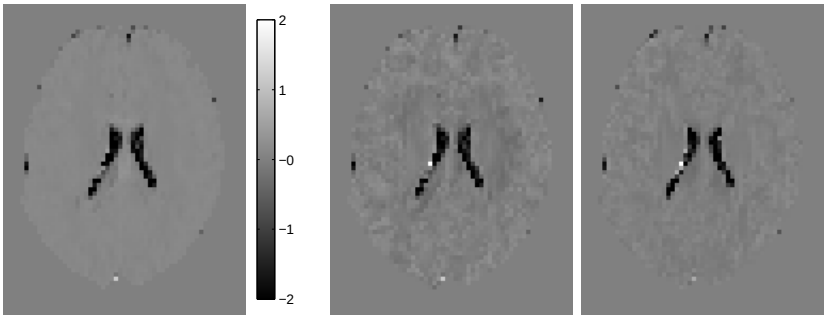


(d) Mean kurtosis (e) Radial kurtosis (f) Kurtosis anisotropy



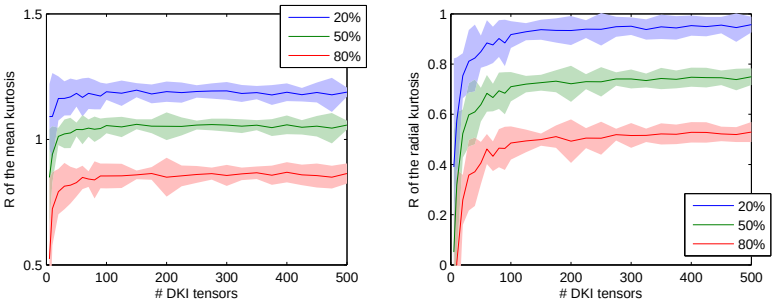
(g) Mean kurtosis (h) Radial kurtosis (i) Kurtosis anisotropy

Figure 7.5: *This figure shows R (Eq. (7.41)) with mean kurtosis, (a),(d),(g), radial kurtosis, (b),(e),(h), and kurtosis anisotropy, (c), (f), (i), of the diffusion kurtosis tensors of a human DKI dataset. The subfigures (a), (b), and (c) are from a dataset recorded with the traditional gradient settings, compared with $O^{b_1}_{2,m}$, $R(\boldsymbol{\theta}, T, O^{b_1}_{2,m})$. In (d), (e), and (f), the gradient settings optimized w.r.t. the mean and radial kurtosis are compared by $R(\boldsymbol{\theta}, O^{b_1}_{2,m}, O^{b_1}_{2,r})$, in terms of the precision of the mean kurtosis, radial kurtosis and kurtosis anisotropy. The subfigures (g), (h), and (i) are from a dataset recorded with the optimized gradient settings $O^{b_1}_{2,m}$, compared with the traditional gradient settings, $R(\boldsymbol{\theta}, T, O^{b_1}_{2,m})$*



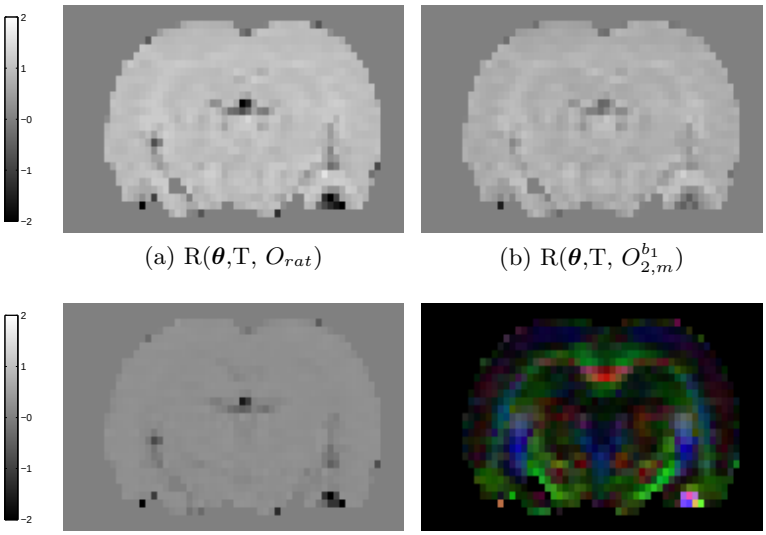
(a) Mean diffusion (b) FA (c) Direction of first eigenvector

Figure 7.6: *This figure shows $R(\boldsymbol{\theta}, T, O^{b_1}_{2,m})$ of the first dataset, for the mean diffusion (a), fractional diffusion anisotropy (FA)(b) and direction of the first eigenvector (c). As is visible in these figures, the precision of the diffusion tensor parameters in the white and grey matter is not substantially changed when the traditional gradient settings are replaced by the gradient settings $O^{b_1}_{2,m}$.*



(a) mean kurtosis (b) radial kurtosis

Figure 7.7: *The subfigures (a) and (b) show the relative performance of the optimized gradient settings as function of the number of elements in the set $\boldsymbol{\Theta}$ that is used during the optimization. All gradient settings were optimized w.r.t. the mean kurtosis. The performance of each optimized settings O is measured with $R(\boldsymbol{\theta}, T, O)$ (Eq. (7.41)) for the mean kurtosis (a) and radial kurtosis (b). For each set of optimized gradient settings, R is computed for $\boldsymbol{\theta}$ of each of the 20k voxels not used in any optimization. The three curves, 20%, 50% (=median) and 80%, give the value of R for which the indicated percentage of tested voxels has a larger value. The shaded areas indicate the 95% confidence interval of a single optimization, obtained by repeating the selection of voxels and optimization 10 times for each number of voxels in the selection.*



(a) R($\boldsymbol{\theta}$,T, O_{rat}) (b) R($\boldsymbol{\theta}$,T, $O^{b_1}_{2,m}$)

(c) R($\boldsymbol{\theta}$,$O^{b_1}_{2,m}$, O_{rat}) (d) FEFA map

Figure 7.8: *The subfigures (a), (b), and (c) show the relative performance of the traditional gradient set T, $O^{b_1}_{2,m}$ and O_{rat} for the mean kurtosis. For localization, an FA map colored with the direction of the first eigenvector (FEFA) is provided in subfigure (d).*

7.A Derivation of a parametric model of the DW-MR images and excess kurtosis

This appendix gives a derivation of the expressions Eq. (7.7) and Eq. (7.30). First, it is shown that the first terms of the Maclaurin series (Taylor expansion around $q = 0$) of the logarithm of the diffusion weighted image intensity $\mathcal{A}(q)$, Eq. (7.6), lead to simple expressions in terms of the diffusion and kurtosis coefficients. Next, the formula that computes the kurtosis in any direction from the diffusion and kurtosis tensors is derived.

The diffusion kurtosis model is based on a truncated Taylor series expansion of Eq. (7.6) around 0. Therefore, we first compute the derivatives of Eq. (7.6).

Fortunately, all derivatives can be computed simultaneously:

$$\left.\frac{\partial^k \mathcal{A}(\boldsymbol{q})}{\partial q^k}\right|_{q=0} = \left.\frac{\partial^k}{\partial q^k} A_0 \int_{x=-\infty}^{\infty} e^{iqx} f_{\boldsymbol{g}}(x) dx\right|_{q=0} \tag{7.42}$$

$$= A_0 \int_{x=-\infty}^{\infty} \left.\frac{\partial^k e^{iqx}}{\partial q^k} f_{\boldsymbol{g}}(x) dx\right|_{q=0} \tag{7.43}$$

$$= A_0 \int_{x=-\infty}^{\infty} \left.(ix)^k e^{iqx} f_{\boldsymbol{g}}(x) dx\right|_{q=0} \tag{7.44}$$

$$= A_0 \int_{x=-\infty}^{\infty} (ix)^k f_{\boldsymbol{g}}(x) dx \tag{7.45}$$

$$= i^k A_0 \int_{x=-\infty}^{\infty} x^k f_{\boldsymbol{g}}(x) dx = i^k A_0 \mathbb{E}_{f_{\boldsymbol{g}}}[x^k], \tag{7.46}$$

where we assume all integrands are finite, the derivatives and integrals exist and are finite and $f_{\boldsymbol{g}}$ is normalized. Note that the odd derivatives are purely imaginary, and thus, even when $f_{\boldsymbol{g}}(x)$ is a-symetric, the magnitude MR images do not change due to this a-symetry.

Since magnitude DW-MR images are recorded, and by construction $\mathcal{A}(q) = \mathcal{A}^*(-q)$, the magnitude is symmetric in q. With Eq. (7.46) and the symmetry, the following expression can be derived for the 2^{nd} derivatives of the logarithm of $\mathcal{A}(q)$:

$$\begin{aligned}\left.\frac{\partial^2 \ln \mathcal{A}}{\partial q^2}\right|_{q=0} &= \left[-\frac{\left(\frac{\partial \mathcal{A}}{\partial q}\right)^2}{\mathcal{A}^2} + \frac{\frac{\partial^2 \mathcal{A}}{\partial q^2}}{\mathcal{A}}\right]_{q=0} \\ &= -\frac{\left(\left[\frac{\partial \mathcal{A}}{\partial q}\right]_{q=0}\right)^2}{A_0^2} + \frac{\left[\frac{\partial^2 \mathcal{A}}{\partial q^2}\right]_{q=0}}{A_0} \\ &= 0 - \mathbb{E}_{f_{\boldsymbol{g}}}[x^2] = -2tD_{\boldsymbol{g}}.\end{aligned} \tag{7.47}$$

Furthermore, the 4^{th} derivative is given by:

$$\begin{aligned}\left.\frac{\partial^4 \ln \mathcal{A}}{\partial q^4}\right|_{q=0} &= \left[-\frac{6(\frac{\partial \mathcal{A}}{\partial q})^4}{\mathcal{A}^4} + \frac{12(\frac{\partial \mathcal{A}}{\partial q})^2 \frac{\partial^2 \mathcal{A}}{\partial q^2}}{\mathcal{A}^3} - \frac{3(\frac{\partial^2 \mathcal{A}}{\partial q^2})^2 + 4\frac{\partial \mathcal{A}}{\partial q}\frac{\partial^3 \mathcal{A}}{\partial q^3}}{\mathcal{A}^2} + \frac{\frac{\partial^4 \mathcal{A}}{\partial q^4}}{\mathcal{A}}\right]_{q=0} \\ &= \left[-\frac{3(\frac{\partial^2 \mathcal{A}}{\partial q^2})^2}{\mathcal{A}^2} + \frac{\frac{\partial^4 \mathcal{A}}{\partial q^4}}{\mathcal{A}}\right]_{q=0} \\ &= -3\mathbb{E}_f[x^2]^2 + \mathbb{E}_f[x^4] = 4t^2 D_{\boldsymbol{g}}^2 K_{\boldsymbol{g}}\end{aligned} \tag{7.48}$$

With these derivatives in terms of the diffusion and kurtosis coefficients, the series expansion of the logarithm of the DW-signal can be given by:

$$\ln \mathcal{A}(q) = \ln A_0 - D_{\boldsymbol{g}} q^2 t + \frac{1}{6} D_{\boldsymbol{g}}^2 K_{\boldsymbol{g}} q^4 t^2 + \mathcal{O}(q^6) \quad . \tag{7.49}$$

In general, the 3D 2^{th} derivative of $\ln A$ can be described by a symmetric tensor of rank 2,

$$\frac{\partial^2 \ln \mathcal{A}}{\partial \boldsymbol{q}^2} = \frac{\partial^2 \ln A}{\partial q_i \partial q_j} = -2t D_{i,j}, \tag{7.50}$$

from which the diffusion in the direction $\boldsymbol{g}$ can be computed by

$$D_{\boldsymbol{g}} = \sum_{i,j=1}^{3} g_i g_j D_{i,j}. \tag{7.51}$$

The 3D 4^{th} derivative of $\ln A$ can be described by a fully symmetric tensor of rank 4,

$$\begin{aligned} \frac{\partial^4 \ln \mathcal{A}}{\partial \boldsymbol{q}^4} &= \frac{\partial^4 \ln \mathcal{A}}{\partial q_i \partial q_j \partial q_k \partial q_l} \\ &= M_{i,j,k,l} = 4t^2 \left(\sum_{i=1}^{3} \frac{D_{ii}}{3} \right)^2 W_{i,j,k,l}. \end{aligned} \tag{7.52}$$

With this 4^{th} derivative and $D_{\boldsymbol{g}}$, the kurtosis in any direction can be computed by:

$$\begin{aligned} K_{\boldsymbol{g}} \equiv K(\boldsymbol{g}) &= \frac{\sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l M_{i,j,k,l}}{4t^2 (\sum_{i,j=1}^{3} g_i g_j D_{i,j})^2} \\ &= \frac{\left(\sum_{i=1}^{3} \frac{D_{ii}}{3} \right)^2 \sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l W_{i,j,k,l}}{\left(\sum_{i,j=1}^{3} g_i g_j D_{i,j} \right)^2}. \end{aligned} \tag{7.53}$$

Note that, due to the division by the diffusion in the direction of $\boldsymbol{g}$, the Kurtosis itself cannot be represented by a rank 4 tensor.

Chapter 8

Conclusions

The goal of this work was to develop new statistical analysis methods for MR images of the brain, focused on functional MRI and diffusion weighted MRI.

Brain images are often acquired with fast imaging techniques, such as Echo Planar Imaging (EPI). Unfortunately, these techniques suffer from distortions. In EPI, a major source of distortions is the inhomogeneity of the main magnetic field. In order to be able to correctly interpret the brain images and to be able to relate the image to the brain anatomy, these distortions need to be minimized. In Chapter 3, a method has been developed with which these field inhomogeneities can be estimated. The field inhomogeneities and other distorting factors were estimated from reference data with a least squares estimator. The results demonstrated that the proposed estimator for the field inhomogeneities is substantially less sensitive to noise in the reference data and has the additional benefit that it returns other interesting parameters, such as a map of the T_2^* time constant of the object.

Multi slice images are generally anisotropic with respect to spatial resolution. That is, the in plane resolution is often much larger than the through-plane resolution. In Chapter 4, a method was developed to combine several multi slice images into a single, isotropic, high resolution image. The acquisition of the MR images was modeled with a set of linear equations and the reconstructed image was a regularized least squares solution of this linear problem. The resulting images demonstrate a substantially improved resolution, compared to any of the original images.

In order to allow a reconstruction with an acceptable computational cost, an advanced method to affinely transform multi dimensional images was de-

veloped. This method avoids aliasing and distortions by splitting the affine transform in a series of carefully designed shear transformations, which can be applied efficiently.

For statistical analysis of MR images, the level of the noise that is present in the MR images needs to be known. In Chapter 5, a Maximum Likelihood (ML) based method was developed with which the noise level of MR images can be estimated from the Rayleigh distributed background mode of MR images. Additionally, a method was developed to automatically select the optimal number of bins of the image histogram that is used for the ML estimate of the noise level. This selection method tries to balance the variance and bias of the noise level estimator. The new ML based method was compared with previously proposed methods and showed a substantial improvement in terms of the root mean squared error.

The most sensitive detection of brain activations from functional MRI datasets requires advanced statistical tests. Advanced likelihood based tests were developed in Chapter 6. These tests allow, in contrast to the general linear model (GLM), direct incorporation of the correlation structure of the noise. Simulation experiments showed that the detection rate of the proposed likelihood ratio test is slightly, but significantly improved compared to the detection rate of the GLM tests. This is true, even when the correlation structure of the noise is estimated separately and the data for the GLM tests is pre-whitened with this estimated correlation structure. Furthermore, it was demonstrated that, for reliable false detection rates with realistic complexity of the noise model, thresholds based on asymptotic theory cannot be used unless the time series contains at least several hundreds of time points for each voxel. Additionally, it was demonstrated that undermodeling of the correlation structure of the noise leads to inferior activation tests.

An important limitation of Diffusion Kurtosis Imaging (DKI) is the sensitivity to noise in the MR images. Due to the extended model of DKI, this sensitivity is increased compared to the Diffusion Tensor Imaging (DTI) model. By optimizing the magnitude and direction of the diffusion weighting gradients of the MR images, the precision of the DKI parameters is maximized. In Chapter 7, a method was developed that optimizes the diffusion weighting gradient settings by minimizing the Cramér Rao Lower Bound (CRLB), which is a lower bound of the variance, of DKI parameters. The improvement in precision that is obtained by optimizing the diffusion weighting gradient settings is substantial; the CRLB of the mean kurtosis in grey matter voxels was improved by a factor 2.1. The optimization of the settings requires a prior distribution of DKI parameters. In Chapter 7, it was shown that the elements selected for the prior distribution substantially influence the obtained precision. For optimal precision of the DKI parameters in brain tissue, all brain tissue types need to be present in the prior distribution. However, when parameters of non brain voxels are included, they will deteriorate the precision with which DKI parameters inside the brain can be estimated. Therefore, DKI parameters from non

brain voxels, such as those of the Cerebrospinal fluid, should not be included in the prior distribution with which the gradient settings are optimized.

Finally, the different methods that have been developed in this thesis are presented together since they can, and in some cases should, be combined for improved statistical analysis of MR images. Even though the actual combinations of methods have not been presented in this thesis, due to practical limitations and time constraints, several relevant combinations that should be investigated further are:

Correct field inhomogeneity distortions in DWI and fMRI
The DWI images recorded for DKI experiments and the BOLD images for fMRI experiments are usually acquired with EPI readout. These images (might) need correction for field inhomogeneities in order to allow correct processing and interpretation of the results. For example, in order to correctly trace the white matter fibers with tractography, the DWI should be anatomically correct. Also, when the results of DTI of fMRI analysis are combined with other anatomical images, such as a high resolution T_1 or even a CT image, which are not (or substantially less) distorted, anatomical correct localisation is relevant.

Improve the resolution of DWI
For DTI and DKI experiments, it would be very beneficial to reduce partial voluming effects by improving the resolution of the diffusion weighted MR images. Currently, the DWI are usually acquired with isotropic voxels. However, when the resolution of isotropic multi-slice MR images is increased, the SNR is strongly reduced. Thus, the high SNR that is required severely limits the maximum resolution with which the images can be recorded. Therefore, a method that improves the resolution without substantial loss of SNR would be beneficial. Such a method was described in Chapter 4. This method combines several anisotropic multi-slice images in an optimal way and each of these anisotropic images can be acquired with a better SNR than a single isotropic high resolution multi-slice image.

Background noise level estimation for activation detection
Knowledge of the noise level is important for activation detection in fMRI experiments. However, the activation detection method developed in Chapter 6 does not use the noise level estimated from the background mode. The reason for this is additional colored noise that is introduced by non task related 'spontaneous' brain activations and possibly activated default mode networks. In our detection method these contributions are regarded as noise with respect to task related activation detection. Our method does not separate the different noise sources and thus the noise level in the background area, which is mainly due to thermal noise in the MR machine, is not representative of the (colored) noise level in the fMRI

time series. However, a lot of current research focuses on the spontaneous activations and the default mode networks. Since these methods try to separate the different 'noise' contributions, knowledge of the noise level introduced by the MR scanner is relevant for these methods.

Optimal experimental design of MR acquisitions

The basic principle of optimal experimental design, and also how it was used in Chapter 7, is applicable to many MR acquisitions. The formalism of optimizing acquisition settings by minimizing the CRLB of interesting parameters is very powerful. In principle almost all acquisition settings can be optimized, as long as a model of the acquisition that includes the acquisition settings as well as the relevant parameters, is available. For example, it might be possible to automatically optimize acquisition settings (TR, flip angle, inversion time, ...) for a contrast of interest (e.g. grey vs. white matter). Such an optimization method might provide especially large benefits when one is interested in several different contrasts that can be acquired simultaneously, as manual optimization of such acquisitions is especially difficult. Furthermore, on the newly developed human 7T MR scanners it is difficult or even impossible to obtain homogeneous B_1 fields, and therefore flip angle, due to the high frequency and absorption. This presents extra challenges for manual design of the acquisition settings. With automatic design, it might be easier to obtain acquisition settings with which accurate and precise quantitative measurements can be performed.

Appendix A

Software

In this section, a general description of the most important software routines that were developed during the authors' PhD project is given. The exact input and output of the routines that are described is well documented in the help text of each routine. Therefore, that information will not be repeated in this appendix.

A.1 Fieldmap estimation

The core routine of the estimation of the main magnetic field inhomogeneities is `LScomputeFieldInhomogenityPoint`. This routine computes the optimal parameter vector of the model Eq. (3.8), as implicitly given by Eq. (3.11). This routine also includes the initialization by the method described in Appendix 3.A. A simple method to compute correction factors for EPI images, based on the parameters estimated for each voxel, is called `calcEPIcorrections`. The routine that can be used to apply the correction factors to an image is called `EPIreconstruct`. Note that these last two routines were developed quickly to visualize the correction; they are not intended as general reconstruction method. Furthermore, it is certainly possible to improve the reconstruction based on the computed field map and other parameters by extending the corrections to more than just the even/odd k-space line phase difference and shift.

A.2 Super-Resolution method for Multi-Slice MRI

For the Super-Resolution reconstruction described in Chapter 4 several routines were developed. First, the matrix vector multiplications $\boldsymbol{X}^T\boldsymbol{S}$ and $\boldsymbol{X}\boldsymbol{o}$ can be computed by `TomographicMRI_Mul` and `TomographicMRI_Transf`. The first of these routines explicitly computes $\boldsymbol{X}$, which it is able to return as output, and is partly implemented in C: `computeVc`. The second routine, `TomographicMRI_Transf`, uses the SSH transformation method for the computation of the matrix vector products. This routine requires an initial 'plan' stage in which the MR acquisition problem is parsed. This parsing consists of two steps. First the affine transformations are decomposed in sets of shear transformations by `planAffineTransform`. Next, the acquisition filter w is incorporated in these transforms. The regularized least squares conjugated gradient solver has been implemented in `cgiterLS`. The regularization matrix, which computes the power in the second derivative of $\boldsymbol{o}$, was created with `createRegularisationmat` and was provided as sparse matrix to the `cgiterLS` routine.

Connected to both the Fieldmap estimation method, Chapter 3, and the Super-Resolution method, Chapter 4, an advanced reconstruction method was developed for the ISMRM reconstruction challenge 2010. This method uses the regularized least squares conjugated gradient solver and a simulation of the MR acquisition process, `simu_MRI`. This simulation routine simulates the acquisition of the k-space samples of each of the potentially multiple coils, with unique spatial sensitivity and phase offsets, when field inhomogeneities and T_2^* decay are present. When a suitable GPU is present and the *GPUmat* toolbox is loaded, the matrix operations are performed on the GPU. To allow the iterative reconstruction with `cgiterLS`, an interface for `simu_MRI` is provided by `MRrecfun`, which also allows the distribution of the computations over multiple worker processes.

A.3 Noise level estimation

In chapter 5 a method was developed to estimate the noise level from the background of MR images. The main routine for this noise level estimation is called `RayleighBackgrndNoiseLvlEst`. This routine uses a histogram of the image, and automatically estimates the noise level.

After the publication of our method we improved our original method to use some spatial information present in the images. Therefore, this method computes the noise level from an image, instead of from the histogram. This new method is called `rayleighBackgroundNoiseLvlMasked`. The improvement is obtained by removing a few voxels around the object's edges, as the true noise free magnitude of these voxels is likely to be unequal to zero (due to Gibbs ringing and/or blurring and/or motion artefacts).

A.4 Functional MRI

The routine that computes the maximum likelihood estimates for the functional MRI data is called `findparmllh_multi`. This function optimizes the likelihood `mllh_rc_multi` of the magnitude and noise model. The experiments for an entire dataset was performed with `testdata_multi`, which was called from the overall test routine `ActivationSensitivityTest` that sets up the simulation experiments.

A.5 Diffusion weighted MR imaging

In preparation for the optimization of the diffusion weighting gradient settings for DKI experiments some general routines for the estimation of both DTI and DKI parameters were developed. The diffusion (kurtosis) tensors were estimated with the 4 routines `ComputeDT[Kurtosis][_parfor]`, potentially with a parfor loop. The 'parfor' command is the command that allows MATLAB to issue the iterations of the loop to different processors (when a 'matlabpool' has been opened). The results of the parfor version and the non parfor version are not numerically exactly equal, as the optimization is evaluated slightly differently in the parfor version. All these routines call functions that calculate the log likelihood function (and gradient) `DifusionTensorLL[Kurtosis]`, which is based on the magnitude predicted by the signal model `Difusion[Kurtosis]Tensor_Apred` and the Rician PDF `logricepdf`.

Several routines to compute relevant parameters from the estimated diffusion (kurtosis) parameters were developed as well. The diffusion tensor eigenvectors and eigenvalues can be computed with `DT_eig`, and the FA with `DT_fa`. The diffusion and kurtosis coefficient in specific directions is returned by `DT_evaluate_Diff_Kurt`, the mean diffusion and kurtosis is returned by `DT_evaluate_mean_Diff_Kurt`, and finally, the radial diffusion and kurtosis is returned by `DT_evaluate_Diff_Kurt_radial`.

A.6 Optimization of diffusion weighted gradients

In Chapter 7 a method was developed to optimize the diffusion weighting settings for DKI analysis. In this section it is explained how the implementation of this method can be applied. The Cramér Rao Lower Bound (CRLB) of the diffusion and kurtosis model parameters is evaluated with `CramerRaoDTKurtosis`. The simulated annealing optimization of the CRLB is implemented in the routine `optimizeDKIgradients` This routine is easy use, with the aid of the provided help text. However, it requires an a-priory set of DKI (or DTI) tensors, which need to be carefully created or selected from a recorded DKI dataset. Therefore, the routine `optgrad_DKIdataset` was created to do just that. The routine `optgrad_DKIdataset` has many options to load processed DKI datasets

(`processDKIdataset`) and extract DKI parameters for the optimization from these results. After the preparation stage that includes selection of the tensors, `optgrad_DKIdataset` starts a distributed computing job in which each task performs an optimization of DKI diffusion weighting gradient settings (`optimizeDKIgradients`). After the optimization, the optimized gradient settings are stored and the CRLB of many DTI and DKI parameters is computed and stored for further analysis. The optimizations of the diffusion weighting gradient settings of which the results were presented in Chapter 7 are started with a script `Optimize_DKIgradients_withdataset_2`. This script calls `optgrad_DKIdataset` with the right parameters.

A.7 General routines

In this thesis, several general routines to support the analysis and presentation of the results were developed:

`progressbar` to indicate the progress of computations, also when computing inside MATLAB workers.

`DistTest` to test if the outcome of some (simulation) process is distributed according to a specific distribution, such as normal with unspecified mean and variance.

`fftresample` to up or down-sample N-D images by zero filling or extraction in the Fourier domain.

`affineTransform` to apply an affine transformation to an N-D image. All methods described in subsection 4.2.7 are supported.

`sphere_best` to evenly distribute sets of points on a sphere.

`imagebrowse` for quick visual inspection of N-D images.

`ploterror` to clearly present (overlapping) uncertainty bounds in figures.

List of Figures

List of Tables

List of symbols

Below is a list of symbols. The numbers, which are links in the pdf version of this manuscript, refer to the equations where the symbol is first used (in that specific meaning).

A	Magnitude of -rice distributed- MR image	Eq. (2.6)
$\boldsymbol{A}$	matrix constructed from α	Eq. (6.90)
$A(\boldsymbol{q};\boldsymbol{\theta})$	DKI model for the magnitude of the MR images	Eq. (7.10)
$\mathcal{A}$	magnitude of the noise free MR signal after diffusion weighting	Eq. (7.6)
A_0	non diffusion weighted signal intensity in diffusion weighted MRI	Eq. (7.6)
$\mathbf{AR}$	A specific Auto Regressive (AR) model	Eq. (6.39)
B	Number of bins of the image histogram used for the ML noise variance estimate	Eq. (5.9)
b	b-value that specifies the strength of the diffusion weighting	Eq. (7.2)
$\boldsymbol{C}$	The linear combinations of parameters in the hypothesis tests	Eq. (6.28)
$\boldsymbol{c}$	Vector that specifies the translation component of an affine transform, or vector with the test position for the hypothesis tests	Eq. (4.22), Eq. (6.28)
D	diffusion coefficient	Eq. (7.4)
$\boldsymbol{D}$	Diffusion tensor	Eq. (7.8)
d	Measure for the distortion	Eq. (4.28)
$\mathbb{E}$	expectation operator	Eq. (2.11)
e	white noise	Eq. (6.2)
F	Fisher information matrix	Eq. (6.35)
$\boldsymbol{f}$	Vector in the frequency domain	Eq. (4.28)
f	model of the MR signal decay	Eq. (3.8)
$f_{i,K}$	Expected count in histogram	Eq. (5.19)

$f(\boldsymbol{x};t)$	Diffusion PDF	Eq. (7.3)
$f_{\boldsymbol{g}}(x;t)$	Diffusion PDF along the direction $\boldsymbol{g}$	Eq. (7.3)
$\boldsymbol{G}$	magnitude and direction of the magnetic gradient for imaging, or for diffusion weighting	Eq. (3.2), Eq. (7.1)
G	magnitude of the magnetic gradient for diffusion weighting	Eq. (7.1)
$\boldsymbol{g}$	gradient direction (unit length)	Eq. (7.3)
H	objective function in optimization of DKI gradient settings	Eq. (7.38)
$H_\cdot$	hypothesis	Eq. (6.28)
h	Transfer function of spatial frequencies	Eq. (4.28)
I	Fisher information	Eq. (7.37)
I_ν	ν^{th} order modified Bessel function of the first kind	Eq. (2.8)
i	imaginary unit, $\mathrm{i}^2 = -1$	Eq. (3.8)
J_ν	ν^{th} order Bessel function of the first kind	Eq. (2.29)
j	rank of C	Eq. (6.28)
i, j, k, l	(summation) indices, used in many locations and does not have explicit meaning	
$\boldsymbol{K}$	Regularization matrix	Eq. (4.7)
K	kurtosis (true; and evaluated from parameters)	Eq. (7.5), Eq. (7.29)
$K_\perp$	radial kurtosis	Eq. (7.31)
KA	Kurtosis anisotropy	Eq. (7.33)
$\boldsymbol{k}(t)$	position in k-space, as function of time	Eq. (3.1)
L	Likelihood function	Eq. (5.13)
l_i	edges of the bins of the histogram	Eq. (5.9)
$\mathfrak{l}$	maximum order of spherical harmonics	Eq. (7.35)
M	Number of voxels with which the noise model is estimated, or with respect to which the DKI gradient settings are optimized	Eq. (6.25), Eq. (7.38)
MK	Mean kurtosis	Eq. (7.29)
m, n, r	matrix or sample sizes	Eq. (6.1)
m	measure function of the CRLB	Eq. (7.28)
N	number of (diffusion weighted) MR images or number of echos	Eq. (3.13), Eq. (4.5), Eq. (7.37)
N_{Br}	Number of elements, estimated by the method of Brummer, in the background mode of the MR image	Eq. (5.3)

N_B	Total number of elements in the selected bins of the histogram	Eq. (5.9)
$\boldsymbol{n}$	unit length direction vector, so a point on $\mathbb{S}^2$	Eq. (7.29)
n	Number of sample points used to integrate a function on $\mathbb{S}^2$ in Chapter 7. Number of dimensions in Chapter 4	Eq. (7.35), Eq. (4.22)
$O_{\cdot}^{\cdot}$	Optimized sets of diffusion weighting settings	subsection 7.3.4
o	object intended to be reconstructed	Eq. (4.1)
$\boldsymbol{o}$	sampled object, intended to be reconstructed, in matrix representation	Eq. (4.2)
P	Power	Eq. (3.13)
P_d	Detection rate	subsection 6.2.10
P_f	False alarm rate	subsection 6.2.10
p	Probability density function (PDF)	Eq. (6.13), Eq. (7.37)
$\boldsymbol{p}$	set of (approximately isotropic) points in $\mathbb{S}^2$	Eq. (7.35)
$\boldsymbol{Q}$	Set of (optimized) gradients for DKI	Eq. (7.37)
Q	Norm of the residu, mainly used to simplify notation of the PDF	Eq. (6.17)
$\boldsymbol{q}$	q-space vector	Eq. (7.1)
q	Magnitude of the q-space vector	Eq. (7.1)
$\boldsymbol{R}$	Relative performance of 2 sets of diffusion weighting gradient settings	Eq. (7.41)
R	The pseudo inverse of C, i.e. the part corresponding to C of the inverse of C joined with its orthogonal complement	Eq. (6.36)
$\boldsymbol{r}$	Position coordinate	Eq. (3.1)
$\boldsymbol{S}$	set of spherical harmonics	Eq. (7.34)
$\boldsymbol{S}_j$	j^{th} MR image	Eq. (4.2)
$S(t)$	Demodulated MR signal, as function of time	Eq. (3.1)
$\mathbb{S}^2$	Surface of the unit sphere (2D surface of 3D sphere)	Eq. (7.29)
$\boldsymbol{s}$	Fisher score vector	Eq. (6.82)
$\boldsymbol{T}$	Affine transformation matrix	Eq. (4.22)
$\boldsymbol{T}_f$	Affine transformation matrix including offset	Eq. (4.22)
$T_{\cdot}$	test .	Eq. (6.32)

T_1	longitudinal relaxation time (of tissues), i.e. the time constant of the magnetization of the tissue in the direction of the magnetic field	
T_2, T_2^*	transverse relaxation time. Nececarily lower than T_1. T_2^* also includes de-phasing due to local field inhomogeneities and molecular interactions	Eq. (3.3)
T_r	Time between the start of the acquisition of 2 k-space lines in an EPI acquisition	Eq. (3.12)
T	Coordinate transform linking object and MR image	Eq. (4.1)
T	Tesla, unit of magnetic strength	Section 1.1
TR	Repetition time; time between subsequent recordings of the same slice/volume.	
$.^T$	transpose of the matrix/vector .	
t	index of time	Eq. (3.1), Eq. (6.2)
t	(effective) diffusion time	Eq. (7.2)
$\boldsymbol{U}$	Matrix with left singular vectors	Eq. (4.12)
$\boldsymbol{V}$	Chapter 4: Matrix with right singular vectors, Chapter 6: covariance matrix of measurements Chapter 6	Eq. (4.12), Eq. (6.3)
v	colored noise	Eq. (6.1)
$\boldsymbol{v}_i$	the i^{th} eigenvector of the diffusion tensor $\boldsymbol{D}$	Eq. (7.31)
$\boldsymbol{W}$	Inverse covariance matrix for the Least Squares DKI parameter estimate	Eq. (7.25)
$\boldsymbol{W}$	Kurtosis describing tensor	Eq. (7.9)
$\boldsymbol{w}$	Optimal weights for integrating on a sphere	Eq. (7.34)
w	Sampling function	Eq. (4.1)
$\boldsymbol{X}$	regressionmatrix	Eq. (6.1)
$\boldsymbol{X}_j$	Matrix that describes the relation between the object and the j^{th} MR image	Eq. (4.2)
X	General random process	Eq. (2.1)
$\boldsymbol{x}$	3D position coordinate (in object space)	Eq. (4.1)
$\boldsymbol{x}$	3D diffusion distance	Eq. (7.3)
Y_l^m	the m^{th} real valued spherical harmonic of order l	Eq. (7.35)
$\boldsymbol{y}$	fMRI data	Eq. (6.1)
$\tilde{y}$	Complex valued MR image voxel	Eq. (2.5)
$\boldsymbol{y}$	3D coordinate (in MR image space)	Eq. (4.1)
α	AR-coefficients	Eq. (6.2)

Γ	Gamma function ($\Gamma(n+1) = n!$)	Eq. (2.10), Eq. (3.2), Eq. (7.1)
γ	gyromagnetic ratio of the imaged nuclei	Eq. (7.1)
Δ	time separation between the leading edges of the diffusion gradient pulses	Eq. (7.2)
δ	duration of the pulsed diffusion weighting gradients	Eq. (7.1)
ϵ	heaviside function	Eq. (2.8)
$\boldsymbol{\varepsilon}$	residu vector	Eq. (6.10)
$\boldsymbol{\Theta}$	set of parameter vectors, either of multiple fMRI time series, or of DKI tensors with which the gradient settings are optimized	Eq. (6.25), Eq. (7.38)
$\boldsymbol{\theta}$	parameter vector	Eq. (6.1), Eq. (7.13)
$\boldsymbol{\lambda}$	parameter vector of the MR decay	Eq. (3.8)
λ	weight for the regularisation	Eq. (4.7)
λ	likelihood ration	Eq. (6.33)
λ_i	i^{th} eigenvalue of the diffusion tensor, sorted in decreasing order ($\lambda_i > \lambda_{i+1}$),	Table 7.2
λ_K	Test statistic	Eq. (5.18)
λ_K^*	Test statistic	Eq. (5.18)
ν	Order of the modified Bessel function	Eq. (2.10)
ρ	correlation (of v)	Eq. (6.4)
$\rho(\boldsymbol{r})$	Local MR signal intensity, i.e. proton density weighted with the applied contrast	Eq. (3.1)
$\boldsymbol{\Sigma}$	Diagonal matrix with singular values.	Eq. (4.12)
σ	Noise standard deviation	Eq. (2.5), Eq. (6.3)
τ	Combination of all fMRI parameters (σ, α, θ)	Eq. (6.28)
$\Phi(.)$	Function that returns the phase of a complex value	Eq. (3.7)
$\varphi(\boldsymbol{r})$	phase difference between images from even and the odd k-space lines	Eq. (3.4)
χ_i^2	Chi-square distribution with i degrees of freedom	Eq. (6.12) Eq. (5.19)
$\omega(\boldsymbol{r})$	Off resonance frequency, as function of location	Eq. (3.3)
Ω	Excited volume when acquiring a MR image	Eq. (3.1)
$\hat{}$	Estimated value	Eq. (3.11)

List of publications

Journal papers

Below is a list of publications of the author of this thesis.

- **D. H. J. Poot**, A. J. den Dekker, R. Achten, M. Verhoye and J. Sijbers, *"Optimal experimental design for Diffusion Kurtosis Imaging"*, IEEE Transactions on Medical Imaging, Vol. 29, Issue 3, p. 819 - 829, March (2010), Impact factor 4.004

- **D. H. J. Poot**, W. Pintjens, M. Verhoye, A. Van der Linden and J. Sijbers, *"Improved B0 field map estimation for high field EPI"*, Magnetic Resonance Imaging, Vol. 28, Issue 3, p. 441-450, April (2010), Impact factor 1.87

- W. Van Hecke, J. Sijbers, S. De Backer, **D. H. J. Poot**, P. M. Parizel and A. Leemans, *"On the construction of a ground truth framework for evaluating voxel-based diffusion tensor MRI analysis methods"*, NeuroImage, Vol. 46, Issue 3, p. 692-707, July (2009), Impact factor 5.694

- A. J. den Dekker, **D. H. J. Poot**, R. Bos and J. Sijbers, *"Likelihood based hypothesis tests for brain activation detection from MRI data disturbed by colored noise: a simulation study"*, IEEE Transactions on Medical Imaging, Vol. 28, Issue 2, p. 287-296, February (2009), Impact factor 4.004

- 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"*, Physics in Medicine and Biology, Vol. 52, Issue 5, p. 1335-1348, March (2007), Impact factor 2.683

- M. W. A. Caan, H. G. Khedoe, **D. H. J. Poot**, A. J. den Dekker, S. D. Olabarriaga, C. A. Grimbergen, L. J. van Vliet, F. M. Vos, *"Estimation of diffusion properties in crossing fiber bundles"*, IEEE Transactions on Medical Imaging, Accepted

Conference proceedings (full paper)

- **D. H. J. Poot**, J. Sijbers, and V. Van Meir, *"General and Efficient Super-Resolution method for Multi-Slice MRI"*, MICCAI 2010, Accepted
- M. W. A. Caan, H. G. Khedoe, **D. H. J. Poot**, A. J. den Dekker, L. J. van Vliet, F. M. Vos, *"Adaptive noise filtering for accurate and precise diffusion estimation in fiber crossings"*, MICCAI 2010, Accepted
- **D. H. J. Poot**, J. Sijbers and A. J. den Dekker, *"An exploration of spatial similarities in temporal noise spectra in fMRI measurements"*, Proceedings of SPIE Medical Imaging 2008, Ed: Joseph M. Reinhardt, Josien P. W. Pluim, Vol. 6914, p. 69142, San Diego, CA, USA, February, (2008)
- W. Pintjens, **D. H. J. Poot**, M. Verhoye, A. Van der Linden and J. Sijbers, *"Susceptibility correction for improved tractography using high field DT-EPI"*, Proceedings of SPIE Medical Imaging, Ed: Reinhardt, Joseph M.; Pluim, Josien P. W, Vol. 6914, San Diego, USA, February, (2008)
- **D. H. J. Poot**, J. Sijbers, A. J. den Dekker and W. Pintjens, *"Automatic estimation of the noise variance from the histogram of a magnetic resonance image"*, IEEE/EBMS Benelux Symposium proceedings, p. 135-138, December, (2006)
- **D. H. J. Poot**, J. Sijbers, A. J. den Dekker and R. Bos, *"Estimation of the noise variance from the background histogram mode of an MR image"*, Proceedings of SPS-DARTS 2006 (The second annual IEEE BENELUX/DSP Valley Signal Processing Symposium), p. 159-162, Antwerp, Belgium, March, (2006)
- J. Sijbers, A. J. den Dekker, **D. H. J. Poot**, R. Bos, M. Verhoye, N. Van Camp and A. Van der Linden, *"Robust estimation of the noise variance from background MR data"*, Proceedings of SPIE Medical Imaging: Image Processing, Ed: Joseph M. Reinhardt and Josien P. W. Pluim, Vol. 6144, p. 2018-2028, San Diego, CA, USA, February, (2006)

Conference proceedings (abstract only)

- J. 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"*, ISMRM2010 proceedings, Vol. 2010, p. 1687, April, (2010)
- J. 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", ISMRM Chapter Benelux conference proceedings, (2010)

- **D. H. J. Poot**, A. J. den Dekker, M. Verhoye, I. Blockx, J. Van Audekerke, A. Van der Linden and J. Sijbers, *"Optimizing the Diffusion Weighting Gradients for Diffusion-Kurtosis Imaging"*, ISMRM2009 proceedings, Vol. 2009, p. 1394, April, (2009)

- **D. H. J. Poot**, A. J. den Dekker and J. Sijbers, *"Pearson Set of Distributions as Improved Signal Model for Diffusion Kurtosis Imaging"*, ISMRM conference proceedings, p. 1383, April, (2009)

- I. Blockx, M. Verhoye, G. De Groof, J. Van Audekerke, K. Raber, **D. H. J. Poot**, J. Sijbers, S. von Horsten and A. Van der Linden, *"Diffusion Kurtosis Imaging (DKI) reveals an early phenotype (P30) in a transgenic rat model for Huntington's disease"*, ISMRM conference proceedings, Vol. 2009, p. 359, April, (2009)

- R. D. Palacios, M. Verhoye, J. Van Audekerke, **D. H. J. Poot**, J. Sijbers, O. Wiborg and A. Van der Linden, *"DKI visualizes hippocampal alterations in the chronic mild stress ratmodel"*, ISMRM conference proceedings, Vol. 2009, p. 744, April, (2009)

- **D. H. J. Poot**, J. Sijbers and A. J. den Dekker, *"Optimizing the Diffusion Kurtosis imaging acquisition"*, European Society for Magnetic Resonance in Medicine and Biology, Valencia, Spain, October, (2008)

- W. Pintjens, **D. H. J. Poot**, M. Verhoye, A. Van der Linden and J. Sijbers, *"Improved EPI Correction: Upgrading An Ultrafast Imaging Technique"*, Liege Image Days 2008: Medical Imaging, March, (2008)

- **D. H. J. Poot**, J. Sijbers, A. J. den Dekker and R. Bos, *"Estimation of the noise variance from the background histogram mode of an MR image"*, Proceedings of the 25th Benelux Meeting on Systems and Control, Heeze, The Netherlands, March, (2006)

Bibliography

[1] W. J. S. Krieg. *Functional neuroanatomy.* Brain Books. Ill Evanston, 1996.

[2] A. K. Afifi and R. A. Bergman. *Functional Neuroanatomy: Text and Atlas.* McGraw-Hill Professional Publishing, 1997.

[3] K. J. Worsley and K. J. Friston. Analysis of fMRI time-series revisited – again. *NeuroImage*, 2:173–181, 1995.

[4] M. S. Cohen. Real-time functional magnetic resonance imaging. *Methods*, 25:201–220, 2001.

[5] R. Goebel, A. Roebroecka, D. Kimb, and E. Formisanoa. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and granger causality mapping. *Magnetic Resonance Imaging*, 21:1251–1261, 2003.

[6] S. Mori. *Introduction to Diffusion Tensor Imaging.* Elsevier, 2007.

[7] D. Weishaupt. *How does MRI work?: An Introduction to the Physics and Function of Magnetic Resonance Imaging.* Springer, Madison, Wisconsin, 2008.

[8] E. L. Barbier, S. Marrett, A. Danek, A. Vortmeyer, P. van Gelderen, J. Duyn, P. Bandettini, J. Grafman, and A. P. Koretsky. Imaging cortical anatomy by high-resolution MR at 3.0T: Detection of the stripe of gennari in visual area 17. *Magnetic Resonance in Medicine*, 48:735–738, 2002.

[9] J. M. Wild, N. Woodhouse, M. N. J. Paley, S. Fichele, Z. Said, L. Kasuboski, and E. J. R. van Beek. Comparison between 2D and 3D gradient-echo sequences for MRI of human lung ventilation with hyperpolarized ^{3}He. *Magnetic Resonance in Medicine*, 52(3):673–678, August 2004.

[10] R. M. Henkelman. Measurement of signal intensities in the presence of noise in MR images. *Medical Physics*, 12(2):232–233, 1985.

[11] H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34:910–914, 1995.

[12] L. Kaufman, D. M. Kramer, L. E. Crooks, and D. A. Ortendahl. Measuring signal-to-noise ratios in MR imaging. *Radiology*, 173:265–267, 1989.

[13] G. McGibney and M. R. Smith. An unbiased signal-to-noise ratio measure for magnetic resonance images. *Medical Physics*, 20(4):1077–1078, 1993.

[14] J. Sijbers, A. J. den Dekker, E. Raman, and D. Van Dyck. Parameter estimation from magnitude MR images. *International Journal of Imaging Systems and Technology*, 10(2):109–114, 1999.

[15] A. van den Bos. *Parameter Estimation for Scientists and Engineers*. Wiley, August 2007.

[16] P. Mansfield. Multi-planar image formation using NMR spin echoes. *Journal of Physics C*, 10:L55–L58, 1977.

[17] S. Mori and P.C.M. van Zijl. Fiber tracking: principles and strategies - a technical review. *NMR in Biomedicine*, 15:468–480, 2002.

[18] P.J. Basser and D.K. Jones. Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. *NMR in Biomedicine*, 15:456–467, 2002.

[19] R. Gruetter. Localized in vivo adjustment of all first-and second-order shim coils. *Magnetic Resonance Imaging*, 20:801–811, 1993.

[20] Y. Zhao, A.W. Anderson, and J.C. Gore. Computer simulation studies of the effects of dynamic shimming on susceptibility artifacts in EPI at high field. *Journal of Magnetic Resonance Imaging*, 173:10–33, 2005.

[21] G. Tao, R. He, A. H. Poonawalla, and P.A. Narayana. Nonlinear image registration with voxel adaptive regularization to correct for EPI-induced geometric distortions. In *Proc. ISMRM*, page 986, 2007.

[22] P. Jezzard and R.S. Balaban. Correction for geometric distortion in echo planar images from B_0 field variations. *Magnetic Resonance in Medicine*, 35:65–73, 1995.

[23] V.J. Schmithorst, B.J. Dardzinski, and S.K. Holland. Simultaneous correction of ghost and geometric distortion artifacts in EPI using a multiecho reference scan. *IEEE Transactions on Medical Imaging*, 20(6):535–539, 2001.

[24] N. Chen and A.M. Wyrwicz. Correction for EPI distortions using multi-echo gradient-echo imaging. *Magnetic Resonance in Medicine*, 41:1206–1213, 1999.

[25] M. Zaitsev, J. Hennig, and O. Speck. Point spread function mapping with parallel imaging techniques and high acceleration factors: Fast, robust, and flexible method for echo-planar imaging distortion correction. *Magnetic Resonance in Medicine*, 52:1156–1166, 2004.

[26] M. Robson, J. Gore, and R. Constable. Measurement of the point spread function in MRI using constant time imaging. *Magnetic Resonance in Medicine*, 38:733–740, 1997.

[27] E.M. Haacke, R.W. Brown, M.R. Thompson, and R. Venkatesan. *Magnetic Resonance Imaging – Physical Principles and Sequence Design*. John Wiley & Sons, 1999.

[28] C. B. Ahn and Z. H. Cho. A new phase correction method in NMR imaging based on autocorrelation and histogram analysis. *IEEE Transactions on Medical Imaging*, MI-6:32–36, 1987.

[29] E. Fieremans, S. Delputte, K. Deblaere, Y. De Deene, B. Truyens, Y. D'Asseler, E. Achten, I. Lemahieu, and R. Van de Walle. A flexible hardware phantom for validation of diffusion imaging sequences. In *Proc. ISMRM*, 2005.

[30] S. Peled and Y. Yeshurun. Superresolution in MRI; application to human white fibre vizualization by diffusion tensor imaging. *Magnetic Resonance in Medicine*, 45:29–35, 2001.

[31] E. Carmi, S. Liu, N. Alon, A. Fiat, and D. Fiat. Resolution enhancement in MRI. *Magnetic Resonance Imaging*, 24:133–154, 2006.

[32] K. Scheffler. Superresolution in MRI? *Magnetic Resonance in Medicine*, 48:408, 2002.

[33] H. Greenspan, G. Oz, N. Kiryati, and S. Peled. MRI inter-slice reconstruction using super-resolution. *Magnetic Resonance Imaging*, 20:437–446, 2002.

[34] R. Z. Shilling, T. Q. Robbie, T. Bailloeul, K. Mewes, R. M. Mersereau, and M. E. Brummer. A super-resolution framework for 3-D high-resolution and high-contrast imaging using 2-D multislice MRI. *IEEE Transactions on Medical Imaging*, 28(5):633–644, May 2009.

[35] K. Malczewski and R. Stasinski. MRI image reconstruction using multiframe integration. In *Proceedings of the Picture Coding Symposium 2007*, Lisboa, November 2007.

[36] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, 2000.

[37] J. A. Fessler and B. P. Sutton. Nonuniform fast fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, 51(2):560–574, February 2003.

[38] K. Fourmont. Non-equispaced fast fourier transforms with applications to tomography. *Journal of Fourier Analysis and Applications*, 9(5):431–450, September 2003.

[39] R. D. Nowak. Wavelet based Rician noise removal for magnetic resonance images. *IEEE Transactions on Image Processing*, 10(8):1408–1419, 1999.

[40] 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 Transactions on Medical Imaging*, 20(1):45–57, 2001.

[41] O. A. Ahmed. New denoising scheme for magnetic resonance spectroscopy signals. *IEEE Transactions on Medical Imaging*, 24(6):809–816, 2005.

[42] Gustavo K. Rohde, Alan S. Barnett, Peter J. Basser, and Carlo Pierpaoli. Estimating intensity variance due to noise in registered images: Applications to diffusion tensor MRI. *NeuroImage*, 26:673–684, 2005.

[43] Marcel Bosc, Fabrice Heitz, Jean-Paul Armspach, Izzie Namer, Daniel Gounot, and Lucien Rumbach. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. *NeuroImage*, 20:643–656, 2003.

[44] Francesco de Pasquale, Piero Barone, Giovanni Sebastiani, and Julian Stander. Bayesian analysis of dynamic magnetic resonance breast images. *Applied Statistics*, 53(3):475–493, 2004.

[45] Levent Sendur, Voichita Maxim, Brandon Whitcher, and Ed Bullmore. Multiple hypothesis mapping of functional MRI data in orthogonal and complex wavelet domains. *IEEE Transactions on Medical Imaging*, 53(9):3413–3426, 2005.

[46] E. R. McVeigh, R. M. Henkelman, and M. J. Bronskil. Noise and filtration in magnetic resonance imaging. *Medical Physics*, 12(5):586–591, 1985.

[47] R. M. Sano. *MRI: Acceptance Testing and Quality Control - The Role of the Clinical Medical Physicist.* Medical Physics Publishing Corporation, Madison, Wisconsin, April 1988.

[48] B. W. Murphy, P. L. Carson, J. H. Ellis, Y. T. Zhang, R. J. Hyde, and T. L. Chenevert. Signal-to-noise measures for magnetic resonance imagers. *Magnetic Resonance Imaging*, 11:425–428, 1993.

[49] J. Sijbers, P. Scheunders, N. Bonnet, D. Van Dyck, and E. Raman. Quantification and improvement of the signal-to-noise ratio in a magnetic resonance image acquisition procedure. *Magnetic Resonance Imaging*, 14(10):1157–1163, 1996.

[50] J. Sijbers, A. J. den Dekker, M. Verhoye, J. Van Audekerke, and D. Van Dyck. Estimation of noise from magnitude MR images. *Magnetic Resonance Imaging*, 16(1):87–90, 1998.

[51] J. P. De Wilde, J. A. Hunt, and K. Straughan. Information in magnetic resonance images: evaluation of signal, noise and contrast. *Medical and Biological Engineering and Computing*, 35:259–265, 1997.

[52] A. J. den Dekker and J. Sijbers. *Advanced Image Processing in Magnetic Resonance Imaging*, volume 27 of *Signal Processing and Communications*, chapter 4: Estimation of signal and noise from MR data, pages 85–143. CRC press, October 2005. ISBN: 0824725425.

[53] M. E. Brummer, R. M. Mersereau, R.L. Eisner, and R.R.J. Lewine. Automatic detection of brain contours in MRI data sets. *IEEE Transactions on Medical Imaging*, 12(2):153–168, 1993.

[54] Lin-Ching Chang, Gustavo K. Rohde, and Carlo Pierpaoli. An automatic method for estimating noise-induced signal variance in magnitude-reconstructed magnetic resonance images. In *SPIE Medical Imaging 2005: Image Processing*, volume 5747, pages 1136–1142, April 2005.

[55] G.M.P. van Kempen and L.J. van Vliet. The influence of the background estimation on the superresolution properties of non-linear image restoration algorithms. In D. Cabib, C.J. Cogswell, J.A. Conchello, J.M. Lerner, and T. Wilson, editors, *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing VI: Proceedings of SPIE Progress in Biomedical Optics*, volume 3605, pages 179–189, 1999.

[56] A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, Tokyo, 3^{rd} edition, 1974.

[57] A. van den Bos. *Handbook of Measurement Science*, volume 1, chapter 8: Parameter Estimation, pages 331–377. Edited by P. H. Sydenham, Wiley, Chichester, England, 1982.

[58] C. A. Cocosco, V. Kollokian, R. K. S. Kwan, and A. C. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4):S425, 1997. http://www.bic.mni.mcgill.ca/brainweb/.

[59] K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2:189–210, 1995.

[60] K. J. Worsley, C. H. Liao, J. Aston, V. Petre, G. H. Duncan, F. Morales, and A. C. Evans. A general statistical analysis for fMRI data. *NeuroImage*, 15:1–15, 2002.

[61] SPM5. software package available at http://www.fil.ion.ucl.ac.uk/spm/software/spm5/.

[62] M. W. Woolrich, B. D. Ripley, J. M. Brady, and S. M. Smith. Temporal autocorrelation in univariate linear modelling of fMRI data. *NeuroImage*, 14(6):1370–1386, 2001.

[63] J. L. Marchini and S. M. Smith. On bias in the estimation of autocorrelations for fMRI voxel time-series analysis. *IEEE Transactions on Medical Imaging*, 18(1):83–90, 2003.

[64] F. Y. Nan and R. D. Nowak. Generalized likelihood ratio detection for fMRI using complex data. *IEEE Transactions on Medical Imaging*, 18(4):320–329, 1999.

[65] S. M. Kay. *Fundamentals of Statistical Signal Processing, Volume II Detection Theory*. Prentice Hall PTR, Upper Saddle River, New Jersey, 1998.

[66] M. B. Priestley. *Spectral analysis and time series*. Academic Press, London, 1981.

[67] E. Vandervliet, G. Nagels, A. Heinecke, W. Van Hecke, A. Leemans, J. Sijbers, and P. M. Parizel. On the cause and mechanisms of the negative BOLD response in fMRI. In *23rd Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology*, page 308, Warsaw, Poland, 2006.

[68] G. A. F. Seber and C. J. Wild. *Nonlinear regression*. John Wiley and Sons, New York, 1989.

[69] D. C. Alexander. *Visualization and Processing of Tensor Fields*, chapter Modelling, Fitting and Sampling in Diffusion MRI, pages 3–20. Springer Berlin Heidelberg, March 2009.

[70] D. C. Alexander. Axon radius measurements in vivo from diffusion MRI: a feasibility study. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, October 2007.

[71] H. Lu, J.H. Jensen, A. Ramani, and J.A. Helpern. Three-dimensional characterization of non-gaussian water diffusion in humans using diffusion kurtosis imaging. *NMR in Biomedicine*, 19:236–247, March 2006.

[72] E. Fieremans. *Validation methods for diffusion weighted magnetic resonance imaging in brain white matter*. PhD thesis, Universiteit Gent. Faculteit Ingenieurswetenschappen, 2008.

[73] 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, April 2009.

[74] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume 2. Hafner Publishing Company, New York, second edition, 1967.

[75] J. H. Jensen, Joseph A. Helpern, A. Ramani, H. Lu, and K. Kaczynski. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine*, 53:1432–1440, January 2005.

[76] Oscar Brihuega-Moreno, Frank P. Heese, and Laurance D. Hall. Optimization of diffusion measurements using Cramer-Rao lower bound theory and its application to articular cartilage. *Magnetic Resonance in Medicine*, 50:1069–1076, 2003.

[77] D.K. Jones, M.A. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine*, 42:515–525, 1999.

[78] E. D. Wu, E. S. Hui, M. M. Cheung, and L. Qi. Towards better MR characterisation of neural tissues using directional diffusion kurtosis analysis. *NeuroImage*, 42:122–134, April 2008.

[79] E. A. Rakhmanov, E. B. Saff, and Y. M. Zhou. Minimal discrete energy on the sphere. *Mathematical Research Letters*, pages 647–662, 1994.