UNIVERSIDAD DE VALLADOLID
Depto. de Teoría de la Señal y Comunicaciones e Ing. Telemática
Escuela Técnica Superior de Ingenieros de Telecomunicación

Tesis Doctoral

A NOVEL FRAMEWORK FOR THE STUDY OF NEURAL ARCHITECTURES IN THE HUMAN BRAIN WITH DIFFUSION MRI

AUTOR
Antonio Tristán Vega
Ingeniero de Telecomunicación

DIRECTOR
Santiago Aja Fernández
Doctor Ingeniero de Telecomunicación

3 de julio de 2009

Tesis:	A novel framework for the study of neural architectures in the human brain with diffusion MRI
Autor:	Antonio Tristán Vega
Director:	Santiago Aja Fernández

El tribunal nombrado para juzgar la tesis doctoral arriba citada, compuesto por los doctores:

Presidente:

Vocales:

Secretario:

acuerda otorgarle la calificación de

En Valladolid, a

Resumen

La imagen por resonancia magnética de difusión (MRI de difusión) es una modalidad de imagen relativamente reciente, que ha permitido el estudio de las fibras nerviosas y su conectividad en la sustancia blanca del cerebro humano *in vivo*. Esta modalidad se basa en el análisis de las direcciones de difusión de las moléculas de agua en el interior de los axones neuronales, que a su vez se relacionan con las orientaciones de las fibras nerviosas. Este fenómeno se ha explotado en el campo de la imagen médica desde mediados de los años noventa, cuando aparecieron los primeros trabajos sobre imagen por tensor de difusión (DTI). Esta técnica está basada en la premisa de que la difusión se puede modelar como un proceso gaussiano, lo cual, como se ha referido ampliamente en la literatura, es estrictamente válido sólo para determinadas configuraciones neuronales muy simples.

Con la aparición de nuevo *hardware* y protocolos para la adquisición bien acelerada o bien de bajo ruido, se han hecho posibles nuevas técnicas para el análisis y procesado de imagen MRI de difusión. De entre ellas, la imagen de difusión de alta resolución angular (HARDI), basada en el muestreo denso de la señal de difusión para todas las posibles orientaciones, es de especial interés en esta tesis, ya que permite la caracterización de la difusión de modo preciso. Al mismo tiempo, la gran variedad de posibles escenarios y propiedades de estos tipos de datos origina nuevos problemas ciertamente complejos, dado que los tradicionales fundamentos para el procesado de imágenes MRI de difusión, tales como la distribución tipo Rician del ruido en los datos escaneados, o el modelo gaussiano para la difusión, no son ya válidos. Estos problemas son el centro de recientes y muy intensos esfuerzos de investigación hoy en día, y la motivación para el presente trabajo.

En esta tesis se propone un marco de trabajo completo para manejar esta diversidad, siendo el último objetivo del mismo la descripción de las microarquitecturas neuronales dentro de la sustancia blanca a partir de la caracterización macroscópica de la difusión. Se incluyen herramientas para la caracterización estadística (modelado de señal y ruido) y el acondicionamiento (filtrado estadístico) de las imágenes de difusión. La novedad de los métodos propuestos reside en su compatibilidad con los escenaros tradicionales basados en señales tipo Rician, pero también con sistemas más modernos con múltiples bobinas y esquemas de adquisición en paralelo.

Sin embargo, la contribución más relevante de esta tesis es la inferencia de información probabilística relativa a las poblaciones de fibras nerviosas en la sustancia blanca del cerebro. Las técnicas descritas se basan exclusivamente en datos HARDI, así que se adaptan perfectamente a los protocolos existentes sin la necesidad de ningún tipo de restricción adicional. Por otro lado, uno de los objetivos principales es evitar cualquier tipo de modelo paramétrico para la difusión, por lo que la asunción de gaussianidad queda definitivamente descartada. Se demuestra que el uso de verdadera información probabilística beneficia en gran medida la descripción de las poblaciones de fibras nerviosas, probándose además que los métodos desarrollados mejoran los existentes en la literatura relacionada.

La evaluación cuantitativa del marco de trabajo completo se realiza mediante datos sintéticos y fantomas diseñados específicamente a tal efecto, comparándose los resultados obtenidos con trabajos relacionados de otros autores. Además la metodología propuesta se ilustra con numerosos ejemplos sobre datos reales, obtenidos de varias instituciones en todo el mundo. Las características de estos datos son suficientemente variadas como para derivar conclusiones con validez general acerca del funcionamiento de los métodos que se proponen.

Abstract

Diffusion Magnetic Resonance Imaging (diffusion MRI) is a relatively recent imaging technique, which has allowed the study of nervous fibers and their connectivity in the white matter of the human brain *in vivo*. It is based on the analysis of the directions of diffusion of water molecules inside the neural axons surrounded by myelin coats, which is related to the orientations of neural fibers themselves. This phenomenon has been used in medical imaging since the mid nineties, with the appearance of the first works on Diffusion Tensor Imaging (DTI). This technique is based on the premise that diffusion may be described in terms of a Gaussian process, which, as it has been widely reported, is strictly valid only for very simple neural configurations.

With the advent of new hardware and protocols for the accelerated or noise–reduced acquisition of MRI data sets, new techniques for image analysis and processing are possible. Among them, High Angular Resolution Diffusion Imaging (HARDI), based on the dense sampling of the diffusion signal for all possible orientations, is of special interest in this dissertation, since it allows the accurate characterization of the diffusion process. At the same time, the wide variety of scenarios and features for such data sets raises a number of new challenging problems of paramount importance, since the traditional foundations of processing algorithms for diffusion MRI, such as the Rician nature of noise in scanned data, or the Gaussian model for diffusion, are no longer adequate. These problems are currently a very active research field and the motivation for the present work.

In this thesis, a new statistical and probabilistic framework is proposed to cope with this diversity, whose ultimate goal is the description of complex micro–architectures of neural fibers through the characterization of the macroscopic diffusion process. It comprises specific tools for the statistical characterization (signal modeling and noise description and estimation) and conditioning (denoising by means of statistics–based image filtering) of diffusion data sets provided by MRI scanners. The novelty of the proposed methods relies on their compatibility with standard scenarios based on Rician signals, but also with modern systems like multiple–coils scanners and parallel MRI acquisition schemes, for which each part of the image is acquired in parallel by an independent receiving coil and further combined with all remaining parts to obtain the whole data set.

The center of the work carried out (and the main contribution of the thesis), however, is the inference of probabilistic information relative to fiber populations in the white matter of the human brain. The techniques described are exclusively based on HARDI data sets, so they perfectly fit currently existing acquisition protocols without further restrictions. The proposed algorithms aim to obviate any specific parametric model for diffusion, completely avoiding the Gaussian assumption for the process of diffusion of water molecules. It is demonstrated that the use of true probabilistic information clearly benefits the inference of fiber populations and orientations, and the methods developed highly improve the previous attempts in the related literature.

The quantitative evaluation of the entire framework is carried out over synthetic and phantom data specifically designed for each part, comparing the results to related works in the recent literature. Besides, extensive experiments over real data sets, gathered from several institutions worldwide, are presented to illustrate the work developed. The features of these data are diverse enough to derive general conclusions on the performance of the new methodology.

Acknowledgments

Many people have contributed to the conclusion of this thesis, helping me whenever I needed help, and supporting me whenever I needed support. First of all, I would like to express my gratitude to all the staff at the LPI, especially to my supervisor, Santiago Aja, for guiding me in the right direction, and to all my partners at lab'25 (and to our former neighbors at GSIC/EMIC). Each one of you has made these years worthy and funny.

I would like to very especially thank Doctor Carl-Fredrik Westin for being such a great host during my stay at Boston. An important part of this work was originated at 1249 Boylston St., and I am pretty sure this thesis would have looked very different without his interest and help. Thanks also to all the people at the LMI/SPL, who made the lab a nice place to stay: Demian, Francisco, James, Julien, Kari, Sylvain, Tom. Of course, I have to acknowledge Raúl San José for his help during those three months (and more).

Following with the people from Boston, the discussions with Doctor Yogesh Rathi were extremely valuable to develop a great part of the work in this thesis. Thanks for taking the time to hear my ideas. Doctor Scott Hoge has also helped a lot with some technical parts on this manuscript, sharing his experience and giving very useful hints.

There are some more people who have contributed to the scientific and technical developments in this thesis. I have to acknowledge Doctors Olivier Salvado and David Raffelt, from the Australian eHealth Research Centre–CSIRO ICT Centre, Brisbane (Australia) for their support. Their collaboration and kind interest in our work helped us to focus the research on image filtering, and highly improved the final result.

I am in debt with Iman Aganj for contacting us and reporting some errors in my original developments, and above all for making me put some new thoughts in my own work. The late discussions we have maintained have greatly improved the overall quality of the central part of this thesis.

Finally, I cannot end these lines without recognizing the endless love, patience, and comprehension of my family and friends during all these years. To my lifetime friends from that certain village, either native or adoptive (too many of you to write your names), thank you for making me not the worst. To my friends at the lab, I could never express how lucky I feel to have you by my side.

Para todas aquellas personas que han hecho de escribir estas páginas algo inolvidable. En primer lugar a mi familia, por su paciencia más allá de todo límite. Y sobre todo a quienes han estado a mi lado para contarlo: a Fede, por toneladas de soporte técnico–moral. A Pablo, por muchos buenos consejos. A Lucilio, por tener siempre un punto de vista diferente. A Dani y a Susana, por poner algo de cordura. A Rubén, por ser tan buen público. A Emma, por su optimismo y sus mañana funciona. A Eloy y a Gonzalo, por tantas horas escuchándome. Y a Vero, por todo lo que ha compartido conmigo y porque ella me ha enseñado las cosas más importantes en estos cuatro años.

Contents

1

Introduction

1.1 Motivation

Magnetic Resonance Imaging (MRI) allows to easily identify the anatomical structures of the brain (cortex and nuclei) *in vivo*. However, with this modality the white matter appears as an homogeneous region, which hinders the complex micro architecture and connectivity of the nervous fibers comprised in this tissue. Diffusion MRI is intended to overcome this drawback, taking advantage of the diffusion of water molecules along the myelinated fiber bundles in the white matter. The first experiment designed to quantitatively describe the diffusion in the white matter was intended by Stejskal and Tanner [Ste65]: the conventional sequence of pulses used for MRI is modified applying two consecutive strong sensitizing gradients with opposite phases. The macroscopic effect is a phase shift of the spins of those water molecules which have moved along the direction of such gradients. It translates in an attenuation of the signal measured in the absence of sensitizing gradients. Analyzing such attenuation for each direction, the diffusion may be quantitatively assessed.

Although this experiment was carried out in 1965, the formal characterization and systematization of diffusion imaging based on MRI is more recent, and mainly due to the pioneering works by LeBihan and Basser [Bas94a, Bas94b, Bas96, Bas00a, Bas00b, LeB86]. The free diffusion of water molecules may be seen as a Brownian motion, where the displacement of a particle in a given time is modeled as a multivariate Gaussian random variable. When the water molecules are confined inside the myelin coats of the neural axons, an isotropic Gaussian behavior cannot be assumed, since the particles are not able to go through these barriers, but only along the axial direction of the axons. In these cases, it is widely accepted that the diffusion may be modeled as a non–isotropic Gaussian process if one single principal direction of diffusion is being imaged. Besides, the physics of the problem impose the radial symmetry of the diffusion, so the entire process can be described in terms of the 3×3 covariance matrix of a Gaussian random variable. As such, the covariance matrix is positive definite and symmetric, so it has only 6 degrees of freedom. This matrix is the diffusion tensor, and those techniques oriented to compute and represent it (or parameters derived from it) at each location of a 3D volume are gathered under the denomination of Diffusion Tensor Imaging (DTI). Since the diffusion tensor has only 6 degrees of freedom, it may be determined from 6 independent gradient directions [Wes01], although it is very common to acquire a higher number of directions to improve the robustness to noise of the estimation [Sal05].

In this dissertation, the more general term Diffusion Imaging will be used to refer to those techniques intended to describe water diffusion inside neural axons which are based on a more general model than Gaussian diffusion. DTI may thus be considered as a special case of Diffusion Imaging, although its great importance in clinical practice makes it a relevant imaging modality itself.

Despite the relative novelty of this imaging protocol, DTI has been successfully used for the diagnosis and analysis of a number of neural illnesses and brain injures [Hor02, Sun04, Lim02, Tay04]. The majority of these works are based on the qualitatively inspection of parameters derived from the diffusion tensor, such as the Mean Diffusivity (MD), related to the amount of diffusion, or the Fractional Anisotropy (FA), related to the different capability of diffusion (tissue structuring) in each spatial direction [Wes01]. However, the use of quantitative parameters extracted from DTI data sets on the fiber tracts of interest is gaining interest [Pag05, Smi06, Smi07]. The following diseases have been intensively addressed in the recent literature:

Epilepsy is sometimes due to injures in some regions of the brain. However, in many situations it is very difficult to determine its origin. It has been observed with DTI that the MD increases and the FA decreases in those zones with malformations of the cortex and in their neighboring tissues. Such zones do not seem pathological in other imaging modalities, so DTI shows a unique potential to discern the focus of epilepsy and better define the injured tissues for their surgical resection.

Multiple sclerosis. This is perhaps the most evident application of Diffusion Imaging in the diagnosis of neural diseases. Multiple sclerosis is a neuro–degenerative illness which produces the reduction, or even the complete destruction, of the myelin coat of the neural axons. Without this barrier, the diffusion of water molecules is not confined to neural axons, so the FA diminishes and the MD augments. Moreover, the change in the MD seems to correlate with the severity of the disease.

Brain ischemia is produced by the temporal or permanent reduction of the blood flow in a region of the brain, which in turn originates a damage in the tissues due to the lack of oxygen. In case of a sudden reduction of the blood flow, the term acute ischemia is used. On the contrary, if this process is gradual, the ischemia is called chronic. DTI allows to distinguish between these types of ischemia by the follow–up of the patient; once again, the disease originate a change in the MD and the FA, which in the acute case is permanent but in the chronic case keeps changing in subsequent examinations.

Alzheimer's disease is especially difficult to detect, since it cannot be diagnosed before the decease of the patient. Although it has been associated to a degeneration of the gray matter, meaningful changes in the white matter are suspected to occur as well. The studies in this direction are preliminary, but a reduction in the FA has been detected in patients potentially affected by Alzheimer's disease.

Schizophrenia. The origin of this illness is unknown, although it is usually associated to biological as well as environmental factors. It is commonly believed that schizophrenia is associated with certain alterations of the brain. DTI studies are mainly based on the comparison of quantitative parameters (MD, FA) extracted from tracts of interest of both control subjects and patients. Some promising results have been reported, mostly related to the reduction of the FA in the positive cases [Buc98, Kub07].

Brain tumors. Although conventional MRI may be used to assess the position and extent of brain tumors, DTI reveals very useful additional information. In particular, it allows to distinguish among different parts of the tumor, the tissues infiltrated, and the edema. DTI studies facilitate the classification of brain tumors and their surrounding tissue, and the quantitative analysis of MD or FA helps to find the zones with metastasis. Moreover, determining the direction of fiber bundles surrounding the tumor is useful for the planning of surgical interventions, minimizing the risk of damaging tracts associated to important functions.

Although DTI shows an enormous potential in the study of the anatomy and connectivity of the white matter, it also has a number of limitations. The principal one is the assumption of a Gaussian model of the diffusion. As mentioned above, this model is only accurate in case one single dominant direction is present in the extent of the voxel being imaged: the longitudinal axis of the fiber bundles are aligned with the principal eigenvector of the diffusion tensor, while the other eigenvectors correspond to transverse directions for which the myelin coats prevents water diffusion. As a consequence, the diffusion tensor has one dominant eigenvalue larger than the other two; if it is represented as an ellipsoid whose main axes are the eigenvectors and the dispersion along each axis is the corresponding eigenvalue, such ellipsoid has an elongated shape, being called a prolate tensor.

On the contrary, there is a number of anatomies in the white matter for which two or more fiber bundles cross in divergent directions, such as the cerebellar peduncle, the upper brain in the joint of the cingulum and the corpus callosum, or the peripheral brain where the longitudinal fasciculus crosses the corona radiata [Mor05]. In these cases, there is a high amount of diffusion in two or more principal directions, and the diffusion ellipsoid is a flattened surface. This shape is often referred to in the literature as an oblate tensor; however, it is worth noticing that in fact an oblate tensor indicates only that the tensor model is not valid for this anatomy: it does not represent a nearly isotropic diffusion in the corresponding plane, but instead that two or more fiber bundles are present in this plane. Moreover, there is a high uncertainty in the nature of diffusion for this kind of voxels, since the orientation of the two (or more) principal directions inside the plane is ambiguous.

Although this uncertainty may be solved in some cases by the context (i.e. by observing the behavior of neighboring voxels), it results obvious that quantitative DTI will be negatively affected by these artifacts. In particular, the existence of oblate tensors yields voxels with artificially reduced FA (which, as mentioned above, is very valuable as a diagnostic predictor). The FA may be seen as the normalized variance of the eigenvalues of the diffusion tensor; for the case of an oblate tensor, this variance is artificially reduced, not as a consequence of the deterioration of the myelin coat, but only as an artifact induced by an incorrect choice for the diffusion model. An interesting study on this topic has been carried out in [Sal05]: at each voxel, the profit of using a more general model than Gaussian diffusion is assessed by means of a goodness of fit statistical test. The conclusion is that a very small percentage of voxels may benefit from the consideration of a general model. Nevertheless, the authors stress the following two limitations or their study:

- If the statistical test is not passed (i.e. it concludes that there is no benefit in considering a non–Gaussian model), it does not necessarily mean that a Gaussian process is enough to model the diffusion. Instead, the meaning is that due to noise such model will not better fit the noisy observations than a Gaussian one.

- The study strongly depends on the number of gradient directions. If only 6 directions are acquired, obviously there is not a way to improve the fitting of the model, since all degrees of freedom are accounted. But if more directions are available, a more complex model may be reliably fitted.

An important conclusion of this analysis is that the limitations of DTI become relevant only if they are not hindered by the presence of noise or by a reduced number of gradient directions. In this sense, the advent of better machinery and new scanning protocols has been fundamental to the development of Diffusion Imaging. In particular, the use of stronger magnetic fields in MRI scanners allows to accelerate the acquisition. Moreover, several receiving coils may be used in parallel either to improve the Signal to Noise Ratio (SNR) or to accelerate the acquisition by acquiring different parts of the image simultaneously. This is the principle of Parallel MRI or Parallel Imaging (pMRI) [Hog05, Lar07, Roe90, Wan00]. These new technologies have allowed to notably improve the quality of the acquisition of gradient images, and at the same time to capture tens or even hundreds of gradient directions in clinical times.

The development of new Diffusion Imaging techniques is therefore the consequence of two factors: first, the *technology push* of the emerging acquisition protocols, which allows more sophisticated models for diffusion. Second, once the two pitfalls of poor SNR and a reduced number of gradients have been surpassed, the limitations of DTI become evident, so general models of diffusion are at the same time a new possibility offered by the current technology and a necessity for the improvement of diagnosis capabilities in Diffusion Imaging.

The search for appropriate models of diffusion beyond tensor representations is currently a very active research field. The number of degrees of freedom to represent for these models is far larger than the traditional 6 free components of the diffusion tensor, so a great number of diffusion gradients has to be acquired. One of the first approaches in this sense, known as Diffusion Spectrum Imaging (DSI) [Tuc03, Tuc04, Wed05], aims to completely characterize the diffusion process by means of the sparse sampling for all possible orientations and magnitudes of the diffusion sensitizing gradients. However, this approach requires an excessive acquisition time to achieve adequate results, so this requirement is often relaxed to characterize all possible directions of the gradients for one given magnitude, which is often referred to as High Angular Resolution Diffusion Imaging (HARDI). There is a relatively large number of efforts in this field, ranging from more or less immediate extensions to multi–tensor models [Kre05, Pel06, Ber07a], continuous distributions of tensors based on deconvolution approaches [And05, Des09, Tou07, Tou08] or analytical techniques [Jia07], or generalized tensor models [Des06], to even more general, non–parametric techniques [Jan03, Öza06] like the popular Q–Balls [Tuc03, Tuc04].

These techniques, although quite experimental, have been successfully used to resolve complex architectures of the white matter, including fiber crossing, bending, and kissing. Compared to the statistical analysis in [Sal05], the new imaging protocols make these approaches useful in a wide variety of anatomical structures, so their advantage is clear for most of the voxels in the image. In particular, the ambiguity of oblate tensors is avoided, since HARDI techniques should be able, in the ideal scenario, to resolve each of the principal directions of crossing. This property is especially desirable in the most popular descriptive (but also quantitative) technique in Diffusion Imaging, known as tractography or fiber tracking. This method consists in the tracking of the fiber bundles at each voxel of the image following the direction of maximum diffusion [Bas00b, Fri06, Mor02, Ten02]. From a number of seeding points, this procedure is repeated for each succeeding voxel until an entire fiber tract is traced according to some stop criterion, providing a very reach qualitative representation of the connectivity between different regions of the brain. Additionally, the measurement of voxel–wise or group–wise parameters on these tracts provides valuable quantitative information. It results evident that, in those voxels with oblate or even spherical tensors, there is a high uncertainty in the orientation of fibers, so it is hard to determine which direction should be followed. Moreover, the most usual stop criterion is to terminate the tracking when a voxel of low FA is found, so an oblate tensor may produce a false positive in this criterion. The advantage of the aforementioned models with respect to both limitations is clear. In fact, different HARDI techniques have been already used with success for this task [Ber07b, Des09].

In spite of their usefulness, the new approaches to the quantification of water diffusion carry on a number of challenging problems opposed to their widespread usage. The following issues can be identified among the most important ones:

The acquisition time, although it can be drastically reduced by the means above described, is still a limiting factor. The detailed description of diffusion compels to measure a large number of gradient directions to fit all the degrees of freedom of the model. Obviously, the more gradient directions, the more accurate the model is fitted, so a trade–off between these two factors has to be reached.

Scanning parameters. There is a number of parameters which may be changed in an MRI scanner: the duration of encoding pulses, the separation between them, the Field of View (FOV), the spatial resolution, and so on. In Diffusion Imaging, the number of sensitizing gradients and their strength (magnitude) are of special interest. Although in conventional DTI the magnitude of the gradients can be maintained relatively low with an adequate performance, the most of the HARDI techniques in the literature may benefit from larger gradient strengths [Öza06, Tou08, Tuc04]. Unfortunately, larger gradient magnitudes induce a greater attenuation of the acquired MRI image for the same noise power, which worsens the SNR of the data set. Yet, applying strong sensitizing gradients may bring on undesirable spatial distortions known as *eddy currents*, which are dependent on the particular direction of these gradients [And02, Net04, Nie04].

The characterization of noise in the diffusion data sets is a critical issue. Although conventional MRI is one of the imaging modalities showing the highest SNR, in Diffusion MRI the need to introduce an important attenuation (which is inherent to the nature if this imaging protocol) notably worsens the SNR. In HARDI techniques, the use of larger sensitizing gradients even stresses the problem. Rician statistics have been traditionally used to describe the noise in MRI [Gud95], which has been the keystone for the estimation of the diffusion tensor [Sal05] or more general diffusion models [Cla08]. Also, the Rician framework has been extensively used in noise removal algorithms [AF08c, Des08, MF09]. However, the new pMRI and multiple–coil techniques (very useful to reduce the acquisition time in clinical practice) reconstruct the MRI data sets in a basically different way to conventional MRI, so the traditional models of noise in MRI are at stake [Con97, Die08]. Given the ubiquitousness of the Rician model in conventional Diffusion Imaging, it is evident that the correct modeling of noise in the new MRI protocols is of capital importance.

The limitations of fiber orientation estimators themselves are an important pitfall in the description of water diffusion. As said before, the limit in the acquisition time prevents from sampling the whole space of the sensitizing gradients for all possible magnitudes an orientations. The solution adopted in HARDI is to sample the entire space of orientations for one single magnitude. As a result, the diffusion process is not completely characterized, and each HARDI technique has to deal with this incomplete information by some means. The way each method overcomes the problem is the principal characteristic which distinguishes it from all others, and usually its main source of error.

From the previous discussion, it follows that the generalization of DTI to more complex diffusion models is in any way a solved problem. Instead, a number of open problems may be identified regarding each of the previously described issues. Partial solutions exist for each of them. For example, and with respect to the acquisition time, a number of pMRI algorithms have appeared in the literature [Hog05, Lar07] that allow to divide it by about two or three times. The main problem of these techniques is the worsening of the SNR. Yet, each of them operates in a very different way, so the model of noise in the final reconstructed data set may highly differ depending on the particular algorithm used. Besides, each algorithm may be used with a different set of parameters, which in turn will affect the noise and the spatial distortions of the Diffusion Images. It is easy to understand that a wide variety of scenarios with very different characteristics may appear, and any imaging protocol intended to work on this kind of data will have to deal with this diversity.

1.2 Objectives

The main goal of the thesis is to establish a framework for the study of water diffusion in the white matter of the human brain, contributing the open problems detected in the recent research efforts. Such framework should include Gaussian diffusion models (DTI) as well as more general descriptors (HARDI). As noted before, the characterization of noise statistics in the emerging MRI protocols, as well as the finding of means to palliate its effect in the imaging techniques, are non–trivial problems which have to be accounted in such a framework. Moreover, the solutions given have to apply for the most common scenarios and current imaging protocols, whose statistical properties and imaging parameters will be very diverse. In particular, the following objectives can be highlighted:

1. Statistical characterization and modeling of the noise in the scanned data, beyond the traditional Rician model. The first task is to describe the first order statistics for the most common protocols involving pMRI and multiple–coil scanners. The correct modeling of the probability distribution of noise is of paramount importance for many applications in the context of Diffusion Imaging. Besides, pMRI is usually based on the subsampling of the data to acquire for each receiving coil; taking advantage of the multiplicity of receivers, the original information may be recovered by means of interpolation, reconstruction in the Fourier domain, and others. As a consequence, the propagation of the uncertainty due to noise from the scanned data to the final reconstructed image is not trivial. In particular, spatial variations of the noise power and correlations of the noisy pattern may appear which are generally not an issue in traditional DTI. Although these artifacts have been preliminary described for some protocols, this is not the case for other important reconstruction methods [Lar07]. The aim is therefore to characterize as completely as possible (i.e. for the most used protocols) the aforementioned propagation of noise to the Diffusion Imaging data sets.

2. Analysis of the impact of noise. Like in any other signal processing task, the presence of noise worsens the accuracy of the studies carried out from experimental data. However, the quantification of the actual impact of noise in DTI is still a matter of issue. Some studies can be found in the literature to this respect [Bas00a, Jon04, Bas06], but they share two limitations: first, they are mainly empirical, so the inference of any conclusion is restricted to the kind of data used in the corresponding study. Second, they are centered on conventional imaging protocols, and therefore on the Rician model. The aim in the thesis will be to generalize the study to account for (almost) all possible scenarios; this is only achievable if a theoretical framework is developed to analytically describe the statistics of noise in the diffusion tensor computed from diffusion images. In other words, this study corresponds to the propagation of uncertainty from the output of the scanner to the final result of the imaging process. The reader should note that the study has been restricted to DTI; for more general HARDI techniques, the variety of methods and mathematical foundations is so diverse that a theoretical framework describing the propagation for all of them is not a realistic target. Instead, the study will be carried out empirically for some simple scenarios.

3. Noise removal. The formal characterization of the propagation of uncertainty in the latter scenario is especially interesting, since this is the part which can be controlled with imaging techniques. The analysis described in the first objective allows to describe how the uncertainty is propagated, but not how it could be reduced, since it is centered on the processes performed inside the MRI scanner. Although the strategies for the improvement of the quality in the acquisition of MRI images are a very active research field themselves, they do not correspond to image processing techniques, so they are beyond the focus of the thesis. On the contrary, once the data have been acquired, the rest of the process lies within the scope of image processing. The aim will be to palliate the effect of noise by means of filtering techniques. These methods will be specific for the kind of images dealt with, as opposed to the extended methodology of multi-channel filtering for each gradient image with existing techniques [AF08c, Des08, Man08]. They will take into account the statistical characterization of noise to fit as general scenarios and protocols as possible.

4. Estimation of fiber orientations. Finally, the limitations of the existing generalized models for water diffusion beyond DTI make it very interesting to find new HARDI techniques to overcome them. In particular, the use of probabilistic models similar to the Gaussian model in DTI may be an advantage for a number of tasks such as stochastic tractography [Fri06]. The probability of the displacement of water molecules in a given spatial direction may be associated to the probability of the existence of a fiber bundle in this same direction, so the estimation of probability densities provides information regarding not only fiber populations but also the certainty in their localization. The last objective of the thesis is indeed the development of new strategies for the probabilistic estimation of fiber orientations in the white matter. Once again, the new methods should be applicable to the diversity of possible scenarios and protocols; more concretely, they should be intended to reduce the sensitivity of common HARDI approaches to the number of sensitizing gradients and their strength.

In the previous discussion the analysis and treatment of noisy signals has been given a high importance. Nevertheless, the most ambitious objective is the last one, since it is the one whose fulfillment will allow to better describe the architecture of the white matter, thus comprising the main contribution to the community of Diffusion Imaging. As a final remark, the qualitative and quantitative evaluation of the techniques derived from the accomplishment of each of the previously defined objectives is an additional objective itself, and yet a challenging task in many cases.

1.3 Methodology

The methodology to be followed is obviously conditioned by the objectives to fulfill. There is a strong dependence between the achievement of the low level goals (i.e. the characterization of the diffusion signals provided by the MRI scanner) and the accomplishment of the aims proposed in the top level (the estimation of fiber architectures). According to these dependencies, a bottom–up scheme is an appropriate methodology, which may be divided in the following tasks:

Development of statistical tools for the analysis of MRI signals. The characterization of noise in the data sets being used requires the deployment of a set of mathematical formalisms and tools. These methods are used to accomplish the first particular objective highlighted in the previous Section. As stated before, the wide variety of MRI hardware and reconstruction schemes originates a multitude of possible scenarios. The first task is thus to group them into similar categories that will share a similar characterization. For some of the most recent categories not properly addressed in the literature, new formalisms have to be developed. This is a requirement derived from the main goal of the thesis, i.e. the general validity of the proposed framework.

Development of tools for the study of the upwards propagation of noise. Analogously to the previous discussion, the analysis of the propagation of uncertainty from the output of the MRI scanner to the estimated fiber structures requires specific mathematical formalisms. As stated before, the general validity of these methods is compromised by the great diversity (and complexity) of the existing approaches.

Development of statistical models for the signals of interest in the presence of noise. With the statistical characterization of noise achieved by the fulfillment of the first two objectives, the filtering of MRI signals may be seen as an estimation problem in which the actual, noise–free signal has to be estimated out from a set of noisy measurements. As in any other estimation problem, the use of prior information on the variables to estimate, in the form of a suitable model, may highly improve the accuracy of the solution. The search for this model, which has to be specific for Diffusion Imaging, will lead to a set of mathematical tools comprising the core of the denoising algorithms pursued in the third objective.

This methodology allows to obtain MRI signals for which the noise–induced artifacts have been drastically reduced. It is opposed to those methods jointly considering the problems of noise modeling and fiber architecture estimation based on this model, which have been used in [Cla08, Fil07, Sal05]. Decoupling both problems makes it possible that the subsequent phases (the estimation of fiber populations) show general validity regardless of the underlying technologies used, which is one of the main issues in this work.

Design of the theoretical framework for the probabilistic estimation of fiber orientations. In this stage the aim is to find the connection between the known probabilism of the diffusion of water molecules in the neural axons and the probabilistic description of fiber populations in the white matter.

Development of novel tools for fiber orientation estimation in the context of the previously described framework. The main limitation of HARDI relies on the incomplete characterization of the diffusion process, so it is necessary to overcome this problem with an adequate set of mathematical tools.

Partial validation of each of the previous stages. Contrary to the previous steps, which have to be carried out sequentially due to the dependencies among them, this one has to be paralleled with all the remaining methods. The evaluation of each of the stages requires to isolate it from all others in order to avoid any influence of the processing of lower levels in the performance of top levels. This stage comprises the quantitative evaluation of the methods developed, mostly based on synthetic and phantom data specifically designed at each stage.

Global validation of the whole framework. In this case, the evaluation will be qualitative, since it has to be performed over real data sets. At the same time, the lack of an extended database with all possible kinds of diffusion MRI data and scenarios makes this validation necessarily incomplete. However, testing the whole processing pipeline for representative situations provides highly valuable information of its joint performance.

Software integration. Together with the validation of each stage, the integration of the methods described in general purpose image processing software tools may be done in parallel with the methodology described.

1.4 Materials

The methods developed in this thesis, especially those related to HARDI techniques, are intended to work over a very specific kind of data. As a result, it is difficult to find a wide test–bed for the validation of the methodology introduced. Moreover, the processing framework is intended to be valid for a wide variety of imaging protocols and scenarios, which makes unfeasible the availability of a complete database with all possible case studies. However, the collaboration with different institutions inside and outside of Spain has allowed the disposal of several Diffusion Imaging volumes of diverse kinds. Since the purpose of the acquisition of each data set is rather different, their characteristics are highly variable. To begin with, five data sets of Diffusion Weighted Images (DWI) are available:

Three HARDI data sets provided by the Surgical Planning Laboratory (Brigham and Women's Hospital) at Boston (Massachusetts, USA)[1]. They are SENSE EPI volumes scanned with a 3 Tesla GE system, with slice thickness 1.7mm, 8 baselines (unweighted MRI images) and 51 diffusion weighted images corresponding to 51 independent gradient directions. For the third volume, however, no baseline images are available: only the attenuation signal, computed as the division of each gradient image by the (averaged) baseline, has been provided as a MatLab data file. The b–value (the parameter related to the strength of the diffusion gradients applied) of each volume is different: $b = 700\text{s/mm}^2$, $b = 586\text{s/mm}^2$, and $b = 1000\text{s/mm}^2$. These volumes will be referred to as BWH1, BWH2 and BWH3, respectively.

---

[1] Courtesy of grant number NIH R01 MH074794 (CFW) from the National Institute of Health, USA.

	b (s/mm^2)	Basel.	Grad.	Res. (mm)	Thick. (mm)	Size
BWH1	700	8	51	0.94×0.94	1.7	$256 \times 256 \times 78$
BWH2	586	8	51	0.94×0.94	1.7	$256 \times 256 \times 81$
BWH3	1000	8	51	0.94×0.94	1.7	$256 \times 256 \times 81$
CSIRO1	3000	8	60	2.31×2.31	2.3	$104 \times 104 \times 54$
CDR1	1000	1	15	1×1	3	$256 \times 256 \times 41$

Table 1.1: Summary of the characteristics of the diffusion weighted data sets used for the validation of the methods proposed. They correspond to brain images of informed volunteers.

An additional HARDI data set, kindly provided by Doctors Olivier Salvado and David Raffelt, from the Australian eHealth Research Centre–CSIRO ICT Centre, Brisbane (Australia). It is a conventional (non–parallel) HARDI data set with 8 baselines and 60 gradient directions, and $b = 3000$s/mm^2. Resolution: $2.31 \times 2.31 \times 2.3$mm. It will be referred to as CSIRO1.

A set of diffusion weighted volumes intended for conventional DTI (they are not HARDI volumes) provided by Doctor Sierra at the Centro Diagnóstico Recoletas, Valladolid (Spain)[2]. They have been scanned with a 1.5 Tesla GE system, taking one single baseline and 15 gradient directions with slice thickness 3mm and $b = 1000$s/mm^2. Among these volumes, a particular one, corresponding to the brain of the author, has been chosen for illustrative purposes where needed. It will be referred to as CDR1.

The main characteristics of these data sets are summarized in Table 1.1.

The previously introduced materials are specific for the methods oriented to Diffusion Imaging. However, some of the stages in the methodology may be validated without the need of these specific resources. To that end, the following additional data are available:

An MRI multiple–coil data set of the brain of an informed volunteer, scanned in an 8 coil, GE Signa 1.5T EXCITE 11m4 scanner, with FSE Pulse Sequence; FOV is 20×20cm (matrix size 256×256) and the slice thickness is 5mm. This data set has been provided as well by the Surgical Planning Laboratory (Brigham and Women's Hospital) at Boston (Massachusetts, USA), and therefore will be referred to as BWH4.

An MRI multiple–coil data set of a human brain, made publicly available[3] by the authors of the PULSAR toolbox [Ji07]. It is an eight-channel head array from a 3 Tesla GE scanner, fast spoiled gradient-echo sequence, with FOV 22×22cm (matrix size 256×256), and will be named PULSAR1.

An MRI multiple–coil data set of a human spine, PULSAR2, acquired in a 3 Tesla whole–body GE scanner from a healthy male volunteer using a four–channel spine array with fast spoiled gradient-echo sequence. FOV is 32×32cm and the matrix size 256×256.This data has also been made publicly available by the authors of [Ji07].

---

[2] TEC2007–67073/TCM from the Comisión Interministerial de Ciencia y Tecnología (Spain).

[3] `http://www.ece.tamu.edu/~mrsl/JIMJI_TAMU/pulsarweb/index.htm`

1.5 Outline of this thesis

The structure of this document follows the methodological stages above described. Chapter 2 briefly describes the fundamentals of MRI, Diffusion Imaging, pMRI, DTI and HARDI. It is not intended as an exhaustive analysis of these techniques, but only as the theoretical basis for the rest of the developments in subsequent Chapters. Besides, the underlying physical processes are not reviewed in depth, in favor of an intuitive posing of the basic equations and formalisms to be used along the document. The general description of the processing framework, which is summarized in Fig. 2.6, is also given in this Chapter.

Chapter 3 is devoted to the derivation of the statistical model for the noise in the MRI signals to be used. The classical characterizations based on Rician statistics are reviewed, and their derivation is detailed. This development is generalized to multiple–coil and pMRI signals. In particular, the GRAPPA protocol is analyzed as an important case study, which allows to conclude that a non–central Chi model allows to characterize signals acquired with the most common MRI protocols. This is especially meaningful because the Rician distribution is a particular case of this general distribution, which can therefore be used with general validity. The main contribution in this sense is the precise statement of the non–central Chi statistic, which has been previously suggested in the literature in an heuristic way [Die08]. Besides, the stationarity of the noisy processes across the image is studied.

Chapter 4 studies the impact of noise, distributed following the general non–central Chi model, in the estimation of the diffusion tensor with Least Squares (LS). This is only a case study inside the analysis of the propagation of noise from scanned data to fiber architecture estimation. However, it corresponds to the most extended Diffusion Imaging technique nowadays, so its relevance is fully justified. Contrary to some previous empirical studies in this sense [Bas00a, Bas06, Cla08, Jon04], the main contributions of the analysis carried out in this thesis are, first, the assumption of a non–central Chi model, which generalizes the applicability of the study to a number of modern protocols; second, this study is analytical, so the conclusions inferred do not depend on the use of any particular data set. Consequently, the analysis has general validity over a range of possible scenarios.

Chapter 5 proposes a number of techniques for filtering of diffusion MRI data sets. As opposed to the techniques commonly used in the literature, based on the separate processing of each gradient image with conventional techniques, a specific methodology is described. It is shown that the high correlation among the gradient images can be exploited as an additional source of prior information. These methods are tested quantitatively over synthetic and phantom data, and qualitatively for real volumes, in the case of Rician noise. The study is restricted to this particular scenario due to the lack of appropriate data sets to validate the algorithms for non–central Chi distributions. Nevertheless, in the last part of the Chapter the study is generalized to this distribution.

The remainder of the document is focused on the main topic of the thesis, i.e. the estimation of neural architectures. Chapter 6 establishes the mathematical formalisms relating the mechanisms of water diffusion to the probability densities for the existence of fiber bundles. They are combined with a popular HARDI technique, known as the Funk–Radon transform (FRT) [Tuc03, Tuc04], to introduce a first approach to the probabilistic estimation of fiber populations, the Orientation Probability Density Transform (OPDT). This technique is thoroughly validated in a number of scenarios, including noisy situations. This latter analysis is useful to complement Chapter 4 with an empirical study of the propagation of uncertainty for the estimator considered.

Chapter 7 extends the idea behind the OPDT. It is argued that the FRT may under some circumstances introduce undesired artifacts in the estimation, so it is necessary to derive additional mathematical tools to avoid this weakness of the OPDT. The first contribution in this Chapter is the introduction of a set of tools based on the FRT which are intended to avoid its limitations. These techniques are further used to derive new estimators of fiber populations alternative to the OPDT. These estimators, together with the OPDT, are the most important contribution to the whole framework discussed in this thesis. Their properties are summed up in Table 7.1, which in turn represents a brief summary of the most important contributions presented in this research work.

The whole framework is validated in Chapter 8 by means of the qualitatively assessment of fiber estimations in known brain anatomies. The methods proposed in Chapters 6 and 7 are tested both isolated and plugged together with the other methods developed in previous Chapters, in order to show the potential of the complete framework. Finally, Chapter 9 highlights the conclusions of the work presented, and proposes a number of open lines derived from it, which comprise an interesting starting point for future research.

1.6 Glossary

ACS	Auto calibrating signal	2.4.1
ADC	Apparent diffusion coefficient	6.1
AWGN	Additive white Gaussian noise	3.2
BLUE	Best linear unbiased estimator	4.3
CMS	Composite magnitude signal	2.4.1
cOPDT	Circulation–based orientation probability density transform	7.4
cQ–Balls	Circulation–based Q–Balls	7.3
DFT	Discrete Fourier transform	2.3
DOT	Diffusion orientation transform	2.7
DSI	Diffusion spectrum imaging	2.7
DTI	Diffusion tensor imaging	2.6
DWI	Diffusion weighted images, diffusion weighted imaging	2.5
EPI	Echo–planar imaging	3.5
ET	Execution time	5.4.7
FA	Fractional anisotropy	2.6.1
FOV	Field of view	2.3
FRT	Funk–Radon transform	6.3
g–SMASH	Generalized simultaneous acquisition of spatial harmonics	2.4.1
GA	Generalized anisotropy	5.4.6
GRAPPA	Generalized autocalibrating partially parallel algorithm	2.4.1
HARDI	High angular resolution diffusion imaging	2.7
iDFT	Inverse discrete Fourier transform	2.3
LMMSE	Linear minimum mean squared error	5.2.1
LS	Least squares	2.4.1
MD	Mean diffusivity	2.6.1
MR	Magnetic resonance	2.1
MRI	Magnetic resonance imaging	2.1
NEX	Number of excitations	2.3
NLM	Non–local means	5.2.2
NMR	Nuclear magnetic resonance	2.1
ODF	Orientation distribution function	6.2
OPDF	Orientation probability density function	6.2

OPDT	Orientation probability density transform	6.5
PDF	Probability density function	2.5
PGSE	Pulse gradient spin echo	2.5.1
pMRI	Parallel magnetic resonance imaging, parallel imaging	2.4
pOPDT	Plane–orientation probability density transform	7.5
pQ–Balls	Plane–Q–Balls	7.5
PSNR	Peak signal to noise ratio	5.4.2
RA	Relative anisotropy	2.6.1
RF	Radio–frequency	2.1
SENSE	Sensitivity encoding	2.4.1
SH	Spherical harmonics	6.4
SMASH	Simultaneous acquisition of spatial harmonics	2.4.1
SNR	Signal to noise ratio	2.3
SoS	Sum of squares	2.4
UNLM	Unbiased non–local means	5.2.2
WLS	Weighted least squares	4.3

1.7 Notation

The notation throughout this dissertation has been kept uniform as far as it is possible. However, the wide specter of techniques and topics dealt with makes this task rather challenging. The uniformity in the nomenclature has been sacrificed at some points in favor of the use of standards in the literature. This inhomogeneity is especially noticeable from those Chapters related to statistical characterization of noise (3 to 5) to those focused on the estimation of fiber populations (6 and 7). For each part, the philosophy has been to keep the most widely accepted notation used in the related bibliography, so that the interested reader can easily follow the developments presented, without the need of getting familiarized with new symbolics. Hence, no general guidelines are provided here, but instead the corresponding notation is introduced and explained where needed.

2

Principles of Magnetic Resonance and Diffusion Imaging

In this first Chapter, the principles of Magnetic Resonance Imaging and Diffusion Imaging are briefly reviewed. The concepts here described are necessary to understand the analysis developed later on, as well as the relevance of the work presented. Therefore, this review is only intended to present the framework for subsequent Chapters, and does not attain the complex physical processes involved in Nuclear Magnetic Resonance (NMR) in detail. The first part is focused on the NMR effect, whose basic mechanisms are fundamental to understand the imaging of soft tissues with MRI. The relation between the resonance frequency and the applied magnetic field are used to give evidence of the Fourier relation between the acquired signal in the **k**–space and the signal in the image domain, which is one of the keystones of the statistical analysis carried out in Chapters 3 through 5. Modern Parallel Imaging (pMRI) techniques are studied to complete the descriptive characterization of the acquisition of the **k**–space, which will take its importance on Chapter 3. The rest of this Chapter is focused on Diffusion Imaging, starting from the characterization of the signal measured in the **q**–space related to diffusion propagators. Once again, a Fourier relation may be proved between these two spaces; Diffusion Imaging is presented as a 6D imaging technique, formed by the Cartesian product of the 3D space of the volume to image and the 3D space of diffusion propagators, each of them with a corresponding Fourier transformed domain (the **k**–space and the **q**–space). The definition of the **q**–space is the basis for the developments in Chapters 6 and 7, while the introduction of the aforementioned 6D space relates these two Chapters with Chapters 3 through 5. In the remaining, Diffusion Tensor Imaging (DTI) is briefly reviewed. The limitations of this model are stressed, leading to the analysis of more sophisticated techniques based on High Angular Resolution Diffusion Imaging (HARDI). This discussion emphasizes the importance of the contributions of the present thesis.

2.1 Introduction

Magnetic resonance Imaging (MRI) is based on a phenomenon known as Nuclear Magnetic Resonance (NMR), first described by Bloch [Blo46] and Purcell [Pur46] in 1946[1]. Under the effect of a magnetic field strong enough, atomic nuclei with unpaired protons *rotate* with a frequency depending on the strength of the magnetic field (and the nature of the atom). This is in fact the resonance frequency of the nuclei for the particular magnetic field strength applied, and the atoms are able to absorb energy at this radio–frequency (RF). In other words, a RF pulse can be used to excite the nuclei, which, once the pulse is removed, emit this electromagnetic energy at the resonance frequency.

The use of NMR to image a given tissue requires the localization of the source

---

[1] The information on this Chapter has been retrieved mainly from [Bas96, Jon05, Lar07, Lia00].

of the electromagnetic energy emitted, in order to infer the spatial position of a given spin density (the concept of spin will be reviewed later on), and therefore the local properties of a given tissue. Since the resonance frequency depends on the strength of the magnetic field applied, the spatial resolution is based on the design of a spatial gradient of the magnetic field: different locations are associated to different magnetic field strengths, and thus to different resonance frequencies: *listening* to different frequencies is the same as studying different locations. This principle was used for the first time by Lauterbur in 1973 [Lau73] to obtain a two–dimensional image. This discovery, together with the Fourier relationship between spin densities and NMR signals, proved by Mansfield and Grannell that same year [Man73], constitutes the basis for modern MRI scanners.

MRI has been used for medical purposes since 1980. This imaging modality provides an excellent contrast between the tissues, is non–invasive, and does not require the use of ionizing radiations, which avoids any secondary effects (as far as it is known). These features make MRI very attractive for the clinical practice, with the only drawbacks of its higher cost compared to other modalities (such us ultrasounds) or its relatively high acquisition times. Besides, NMR–derived effects may be used in other MRI modalities: apart from anatomical MRI, functional MRI or diffusion MRI provide complimentary information and are the focus of important research efforts and interest. The focus of this dissertation is on diffusion MRI, but all these modalities share a number of common principles to be discussed in the remaining of this Chapter.

2.2 Physics of Magnetic Resonance Imaging

Protons, neutrons, and electrons show an angular moment known as spin, which may have the values $\pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \pm 2, \pm\frac{5}{2}\dots$ When these particles are paired, their spins are paired as well, so they cancel each other. This is the reason why NMR is only feasible with unpaired protons. In MRI, the particles considered are hydrogen nucleus associated to the concentration of water molecules. In this case, the spin is reduced to values $\pm\frac{1}{2}$. The spin is a property of elemental particles, so it has to be analyzed in the scope of quantum mechanics. However, in MRI, spin systems and not individual spins are analyzed, so their macroscopic behavior may be accurately described with classic magnetic fields theory. In this sense, the spin may be seen as a microscopic magnetization vector originated by the movement of electrons around the nuclei, much like the magnetic vector induced by a round wire conducting an electric current. In the absence of an external stimulus, spins are randomly distributed, so the macroscopic magnetization is $\mathbf{M} = \mathbf{0}$. When an external magnetic field $\mathbf{B}_0$ is applied, the spins align with its direction (by convention, it is assumed to be the z axis). Unlike a classical magnetic dipole, spins may adopt one of two energy states inducing opposite microscopic magnetizations. The low energy state has a slightly greater probability, so an overall magnetization $\mathbf{M}$ appears aligned with $\mathbf{B}_0$. At the same time, the magnetization vector of individual spins is subject to a precession movement, previously referred to in Section 2.1, around $\mathbf{M}$. Its frequency ω_0 is commonly known as the Larmor frequency or the resonance frequency, and may be written as:

$$\omega_0 = \gamma \|\mathbf{B}_0\| = \gamma B_0, \tag{2.1}$$

with γ the gyromagnetic ratio. As previously stated, the Larmor (resonance) frequency depends on the strength of $\mathbf{B}_0$, and on the properties of the tissue being imaged through γ. The phase of the precession movement for each spin is random, so the macroscopic effect is that the component of $\mathbf{M}$ in the transverse (xy) plane is null, while there is a net longitudinal component in the z direction. Once the spins have been aligned with $\mathbf{B}_0$, and are precessing at frequency ω_0, they are able to absorb energy from a radio frequency pulse $\mathbf{B}_1(t)$. This pulse may be thought of as a circularly polarized magnetic field rotating at frequency ω_0 in the plane xy, so it may couple with the precession movement. The first effect is the coherence on the rotation of the spins, which induces a rotating component in the transverse plane xy. Besides, the particle is able to absorb energy; the overall effect is that an effective magnetic field $\mathbf{B}_{\text{eff}}$ appears aligned with one of the directions x or y, the precession of the spin follows this direction and the microscopic magnetization drifts from the direction of $\mathbf{B}_0$ a time–dependent angle α. Controlling the duration τ of the pulse $\mathbf{B}_1(t)$, the final value of α may be fixed. For real–world applications, durations which produce angles $\alpha = 90^o$ or $\alpha = 180^o$ are used, and are commonly known as 90^o or 180^o pulses. The angle of **M** is changed by exciting the spins with electromagnetic energy at the Larmor frequency. When the pulse $\mathbf{B}_1$ is removed, the spins free the energy they have previously absorbed, going back to their initial state aligned with $\mathbf{B}_0$. This process is called relaxation and, during it, the energy is emitted in the form of a RF signal which may be received by an antenna (in MRI, antennas are receiving coils placed in the MRI scanner).

The relaxation of the spins is associated with two different physical processes: the pulse $\mathbf{B}_1(t)$ cedes energy to the spin changing its precession movement; after a time τ, the longitudinal component of **M** is removed. On the other hand, $\mathbf{B}_1(t)$ produces the coherent rotation in the xy plane, yielding a net transverse component. Once the pulse is removed, both components return to their original states. Both processes occur at the same time, an are basically independent, although in general the second one is much faster. Therefore, there exist two relaxation times:

T_1: is the time for the longitudinal component to return to its original state through the emission of electromagnetic energy at Larmor frequency. This is the spin–lattice relaxation, corresponding to the exchange of energy between the spin–system and its surroundings.

T_2: is the time for the transverse component to return to its original state, associated to thermal equilibrium between spins: when the pulse $\mathbf{B}_1(t)$ is applied, the phases of the spins reach a certain coherence, yielding a net transverse component, which disappears when the pulse is removed. This is the spin–spin relaxation.

Both times T_1 and T_2 refer to the time constant of the exponential laws ruling the relaxation processes. In general, $T_1 \gg T_2$. Measuring relaxation times of the longitudinal and transverse components of **M**, different properties of the tissues may be inferred. This is the principle of T_1 and T_2 imaging modalities (see Fig. 2.1). In the present work, the T_2 modality is of especial importance since it is the only one which allows the measurement of diffusion.

2.3 The **k**–space and the **x**–space

The Larmor frequency depends on the strength of the external magnetic field, $\mathbf{B}_0$. This property may be used to infer spatial information by the use of field gradients. A spatial gradient is applied to $\mathbf{B}_0$ in the z direction while the radio frequency pulse $\mathbf{B}_1(t)$ is active. This implies that the Larmor frequency varies for each plane z_p, so only one of the planes is excited by the pulse, being able to absorb electromagnetic energy. This principle is used in MRI to select an image slice:

$$\omega_0(z_p) = \gamma B_0 + \gamma G_z z_p, \tag{2.2}$$

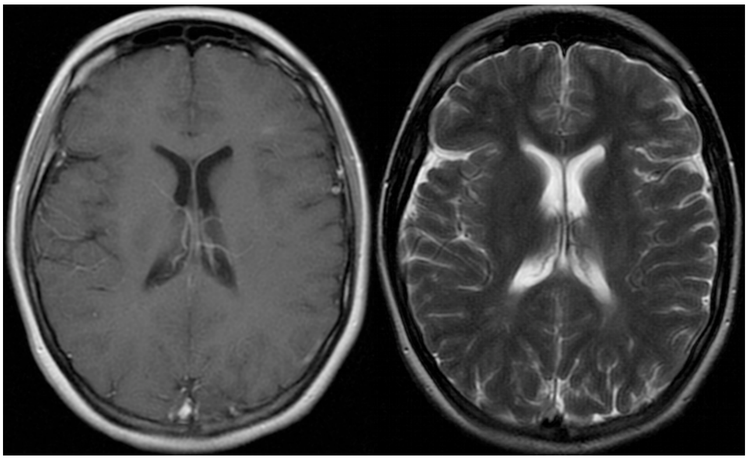


Figure 2.1: Examples of anatomical MRI images of different modalities: T1 (left), and T2 (right). In these modalities, each pixel represents the relaxation time (longitudinal component for T1, and transverse component for T2) after the application of the RF pulse.

where G_z is the modulus of the gradient applied in the slice direction. The spatial encoding for the xy plane is more complex. A combination of gradients in the x and y directions simultaneously may be thought of. With this strategy, for each selected slice z_p, there is a plane defined by the two gradients G_x and G_y:

$$\omega_0(x,y,z_p) = \gamma B_0 + \gamma G_x x + \gamma G_y y + \gamma G_z z_p, \tag{2.3}$$

which defines lines in the angle $\tan^{-1}(G_x/G_y)$ with the same Larmor frequency. The collected signals will be the superposition of the spins along these lines. Varying the ratio G_x/G_y, a projection image similar to computerized axial tomographies may be obtained. The main drawback is the need to infer the spatial information from projections, as is the case with tomographies.

A phase/frequency encoding is used in practice: once the plane z_p has been chosen with G_z, a pulsed gradient G_y is applied, so the Larmor frequency is different for each point along the y axis. If the duration and amplitude of G_y are properly chosen, when G_y is removed the points along y have linearly spaced phases, so that their Larmor frequencies return to their original value but their spins have different phases. Then, a pulsed gradient G_x is applied, varying the Larmor frequency along x. The RF signal is measured while G_x is being applied (see Fig. 2.2).

The x direction is encoded in the frequency of the emitted signal, while the y direction is encoded in its phase. Unfortunately, this scheme is prone to an ambiguity in the phase encoding: the superposition of several signals with different phases has a phase which is a function not only of the phases of the original signals, but it depends on their amplitudes as well. In practice, this means that the acquisition has to be repeated several times for slightly different values of G_y. The resolution in the y axis is given by the number of repetitions used in the acquisition process, while the resolution along x depends on the number of samples taken at each line.

The advantage of this encoding scheme is that it may be proved that the frequency/phase plane is in fact the two–dimensional inverse Fourier transform of the spatial information [Kin84]. Without entering into unnecessary details, note that for each phase encoding the radio frequency signal is the superposition of all the harmonics $\omega_0(x,y_j,z_p) \equiv \omega(x)$ (with y_j the location corresponding to this phase encoding), weighted by the actual value of the energy emitted at location x with Larmor frequency $\omega(x)$. The relation with the Fourier transform in the direction of the y axis is not so trivial, but in general the received radio frequency signal s may

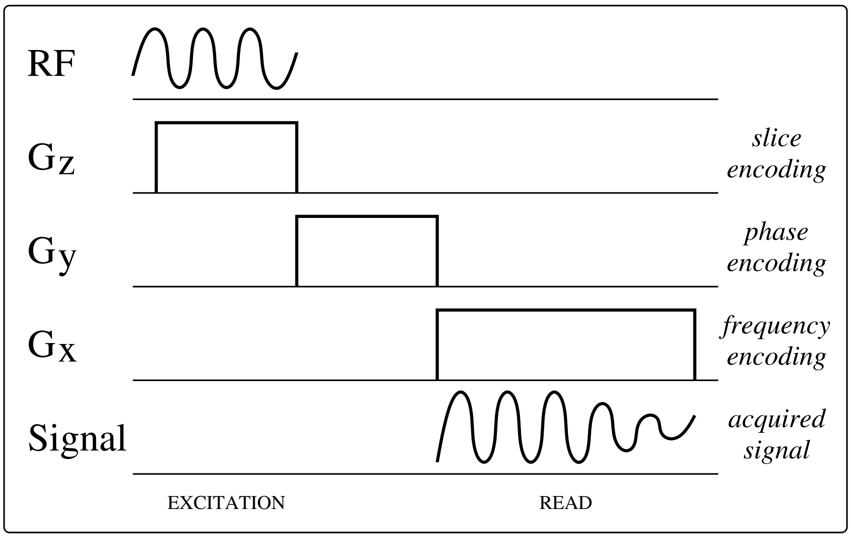


Figure 2.2: Magnetic field gradients for the encoding of spatial information in MRI. The pulse G_z used for the encoding of the slice z_p is applied at the same time as the radio frequency pulse $\mathbf{B}_1(t)$, with a strength greater than G_x and G_y. Therefore, only the spins in the slice z_p are able to absorb the radio frequency energy provided by $\mathbf{B}_1(t)$. G_y alters the frequency of the precession of spins for different y positions. Before G_x is applied, all spins return to the Larmor frequency ω_0, with different phases depending on y. When G_x is applied, the resonance frequency changes in the direction x, so each pair x, y is identified by a unique frequency and phase in the composite RF signal detected by the coils.

be modeled as:

$$s(\mathbf{k}) = \int_V W(\mathbf{x})\rho(\mathbf{x}) \exp\left(j2\pi\mathbf{k}^T\mathbf{x}\right) d\mathbf{x}, \tag{2.4}$$

where $\rho(\mathbf{x})$ is the spin density at spatial location $\mathbf{x}$ within the Field Of View (FOV) of the scanner, V, which is the whole spatial domain for which the tissues are imaged. $W(\mathbf{x})$ accounts for the possibility that the sensitivity of the receiving coil is different for each location. Eq. (2.4) is obviously the (weighted) inverse 2D Fourier transform of $\rho(\mathbf{x})$ in the dual variable $\mathbf{k}$ for each slice z_p. Following this traditional notation, the signal acquired by the receiving coil is said to be in the **k**–space, while the signal of interest, i.e. the spin density, is defined on the image domain, which in this dissertation will be referred to as the "**x**–space"[2].

MRI scanners use the protocol described in Fig. 2.2 to acquire the **k**–space line–by–line (see Fig. 2.4): for each repetition of the phase encoding, a pulsed G_x is applied. The frequency encoded radio frequency signal is sampled to achieve a whole line of the two–dimensional inverse Fourier transform of $\rho(\mathbf{x})$. Then, a two–dimensional Discrete Fourier Transform (DFT) is used to recover the **x**–space from the sampled **k**–space for each slice z_p. The entire acquisition process is often repeated several times, so that multiple samples of each point in the **k**–space are available. The average of all these measurements serves to improve the SNR of the data set. The number of measurements is commonly referred to as the Number of Excitations (NEX).

2.4 Fundamentals of Parallel Imaging

The process described in the previous Section is the principle of MRI, being its main problem the high acquisition time inherent to the image formation process.

---

[2] The more conventional notation '**r**' for the spatial location has been substituted by '**x**', since in Chapters 6 and 7 '**r**' will have another meaning.

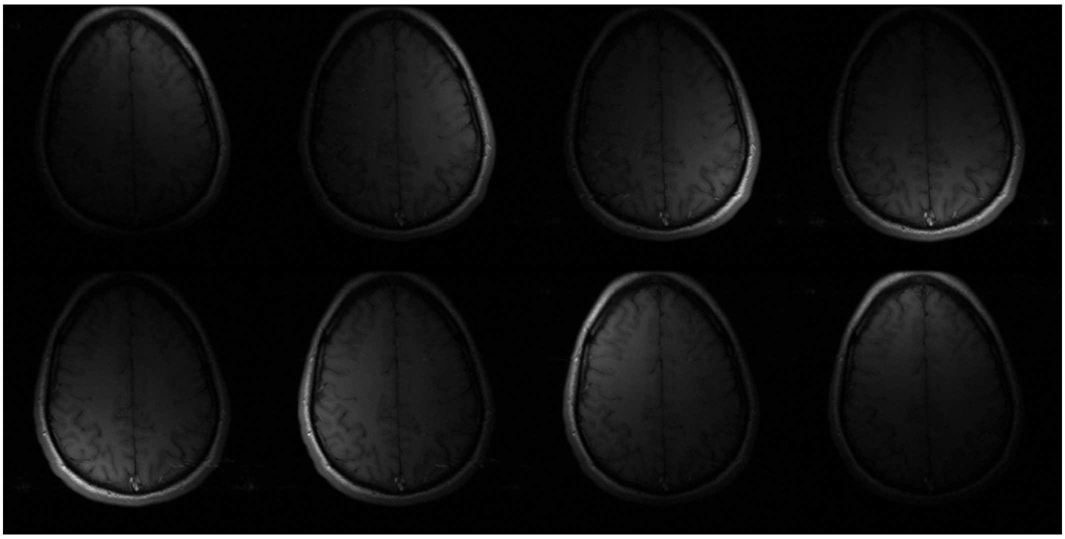


Figure 2.3: An example of a multiple–coil system for Parallel Imaging (BWH4, see Section 1.4). Eight receiving coils have been placed at different spatial positions; since each of them has its own sensitivity, each one is more sensible to a certain subset of the **x**–space. A given antenna receives the radio frequency signal with higher strength for the spatial region for which it is more sensible. This is a kind of spatial encoding, which in Parallel Imaging is exploited to reduce the number of phase encoding repetitions an accelerate the acquisition.

It may be computed for each slice as:

$$T = T_R \cdot N \cdot \mathrm{NEX}, \tag{2.5}$$

where T_R is the repetition time, i.e. the time it takes for the selected plane z_p to return to its equilibrium after it has been excited by the pulses. N is the number of steps used for phase encoding, directly related to the resolution in the y axis. For example, for $T_R = 1.5s$, $N = 120$ and $NEX = 2$, it takes 360 seconds to acquire a single slice. For 20 slices, the acquisition time is more than two hours. In the case of Diffusion Weighted Imaging (DWI), where multiple data sets have to be acquired, acquisition times may become an important limitation.

Acquisition times may be reduced with modern MRI techniques, especially in the case of multiple receiving coils scanners. When a number of independent antennas (receiving coils) work together, each of them acquiring a subset of the **k**–space, Fourier domain information may be retrieved faster. This is the principle of Parallel Imaging, or Parallel Magnetic Resonance Imaging (pMRI). The price to pay when using these techniques is the necessity of the reconstruction of the whole **k**–space (or **x**–space) from its parts. Besides, the traditional statistical characterization for MRI signals, which has been the keystone for most of the image processing techniques for this modality, is no longer valid. This issue will be discussed in Chaper 3.

Multiple array coils in MRI were firstly intended to improve the SNR in MRI acquisitions while maintaining a large FOV. It was in the late 1990s when parallel acquisition schemes started their development. As has been mentioned above, the most time consuming task in MRI is the spatial encoding in the in–plane direction, i.e. the repetitions needed for phase encoding. The idea behind pMRI is to use spatially separated antennas to perform some of the spatial encoding taking advantage of the spatial sensitivity $W(\mathbf{x})$ in eq. (2.4) [Lar07], and so to allow the use of less repetitions for phase encoding. An illustrative example may be found in Fig. 2.3. In its simplest form, the use of multiple coils allows a sub–sampling of the **k**–space, whose most immediate effect is the aliasing in the **x**–space due to the violation of Nyquist criterion (see Fig. 2.4). The signals received by each antenna

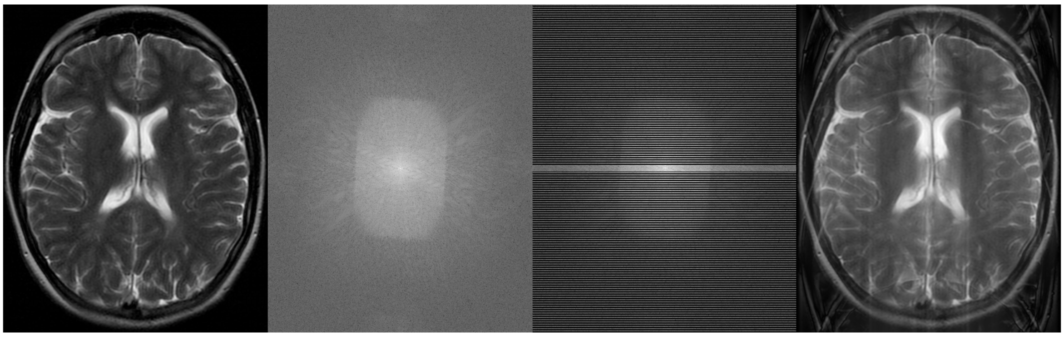


Figure 2.4: Effect of **k**–space sub–sampling. From left to right, the original T2 image (from the synthetic BrainWeb database, see Fig. 2.3); the ideal **k**–space for this image (computed as the DFT of the original image) in logarithmic units; the sub–sampled **k**–space, in logarithmic units, resulting form the elimination of one of each two lines in the y direction; the modulus of the image domain (**x**–space) reconstructed from the sub–sampled **k**–space. The Fourier relation between the **k**–space and the **x**–space explains the aliasing in the image domain if the **k**–space is sub–sampled violating Nyquist criterion. As a result, two points (in this over–simplified scenario) of the original image contribute to each image location of the reconstructed image. Parallel Imaging algorithms eliminate this artifact using the redundant information from several receiving coils.

has to be combined in some way to avoid this artifact. The way the signals are fused is the reconstruction scheme which defines each pMRI algorithm.

The main drawback of pMRI is the reduction of the SNR of the images due to reduced Fourier averaging. For non–parallel schemes, the computation of the DFT produces an averaging of the noise components over the samples of the **k**–space at each point of the **x**–space, notably improving the SNR. With parallel acquisitions, not all the samples in the **k**–space are acquired, so the increase in the SNR is minor. Besides, reconstruction schemes introduce an amplification of the noise in the **x**–space known as the g–factor (where ‘g’ stands for ‘geometric’).

These concepts, together with their implications on DWI, will be thoroughly analyzed in Chapters 3 and 4 respectively, so a detailed study makes no sense at this point. Nonetheless, the different reconstruction algorithms work in different ways, leading to very different properties in the reconstructed image. In the remainder of this Section, the most common Parallel Imaging techniques are reviewed; among them, SENSE and GRAPPA are the most relevant due to their intensive use in commercial MRI applications. Hence, more details are given about these two approaches. For a detailed description on all these algorithms, see [Gri02, Hog05, Lar07, Pru99].

2.4.1 Parallel reconstruction algorithms

Sensitivity Encoding (SENSE)

In the easiest scenario, SENSE works by eliminating one of every two lines in the **k**–space. The acceleration rate is $r = 2$, since only half the repetitions for phase encoding have to be acquired [Pru99]. Assuming there are two receiving coils (the RF signals are acquired simultaneously for both of them), the corresponding signals in the **x**–space will be, see eq. (2.4):

$$\begin{aligned}\widehat{S}_1(x,y) &= W_1(x,y)\rho(x,y)+W_1\left(x,y+\frac{Q}{2}\right)\rho\left(x,y+\frac{Q}{2}\right)\\ \widehat{S}_2(x,y) &= W_2(x,y)\rho(x,y)+W_2\left(x,y+\frac{Q}{2}\right)\rho\left(x,y+\frac{Q}{2}\right),\end{aligned} \tag{2.6}$$

for $\mathbf{x} = [x, y]^T$, and Q the number of lines in the y direction: since the $\mathbf{k}$–space has been decimated with a factor 2 in the phase encoding direction, an aliasing in the $\mathbf{x}$–space is induced, resulting in the overlap of replicated versions of the image (see Fig. 2.4). If the sensitivity of each coil at each image location $\mathbf{x}$ is known (i.e. if $W_l(\mathbf{x})$ may be characterized by means of a calibration process), the linear system in eq. (2.6) may be solved for $\rho(\mathbf{x})$. It is useful to give here some remarks:

- If more than two coils are available, eq. (2.6) turns into an overdetermined system which may be solved by means of Least Squares (LS) techniques to palliate the effect of noise.
- SENSE requires to characterize $W_l(\mathbf{x})$ for all coils l. This implies the need for a calibration stage, being this one of the main drawbacks of this technique.
- Since only half (or in general $1/r$ times) the lines in the $\mathbf{k}$–space are used, Fourier averaging is reduced, and the noise power in the $\mathbf{x}$–space is increased by a factor $\sqrt{r}$. Besides, if LS are used to solve for ρ, an additional space–dependent amplification of noise, the g–factor, is produced which worsens the SNR (see [Lar07] for details).

Finally, note that the reconstruction in SENSE is carried out once the $\mathbf{x}$–space has been recovered via DFT, hence the denomination “image domain method” as opposed to the spectral or hybrid methods presented below.

Simultaneous Acquisition of Spatial Harmonics (SMASH)

SMASH [Sod97] keeps the coil information in the image domain ($\mathbf{x}$–space), but operates on the target data in the $\mathbf{k}$–space. Briefly speaking, SMASH reduces the number of repetitions for phase encoding using the coil sensitivities to explicitly obtain missing lines. Suppose that the receiving coils are placed along the direction y of phase encoding; the linear combination of their sensitivities may be used to obtain a modulation in the y direction, related to the m–th harmonic, as follows:

$$\sum_{l=1}^{L} w_l^m W_l(\mathbf{x}) \simeq \exp\left(j2\pi m \cdot \Delta k_y \cdot y\right), \; \Delta k_y = \frac{2\pi}{Q}, \tag{2.7}$$

where Δk is the minimum step between consecutive lines in the $\mathbf{k}$–space, and the weights w_l^m associated to the m-th harmonic for coil l may be obtained with LS. Once the phase encoding is explicitly performed for the acquired lines, eq. (2.7) may be casted into eq. (2.4) to obtain the relation:

$$\begin{aligned} \hat{s}(\mathbf{k}) &= \int_V \left(\sum_{l=1}^{L} w_l^m W_l(\mathbf{x}) \right) \rho(\mathbf{x}) \exp\left(j2\pi \mathbf{k}^T \mathbf{x}\right) d\mathbf{x} \\ &\simeq \int_V \rho(\mathbf{x}) \exp\left(j2\pi \mathbf{k}^T \mathbf{x}\right) \exp\left(j2\pi m \cdot \Delta k_y \cdot y\right) d\mathbf{x} \\ &= \int_V \rho(\mathbf{x}) \exp\left(j2\pi \left(k_x x + (k_y + m\Delta k_y) y\right)\right) d\mathbf{x} \\ &= s(k_x, k_y + m\Delta k_y), \end{aligned} \tag{2.8}$$

for $\mathbf{k} = [k_x, k_y]^T$. It is clear form eq. (2.8) that SMASH uses an artificial sensitivity (by linear combination of coil sensitivities) to refine the $\mathbf{k}$–space in the y direction, so the phase encoding repetitions may be reduced. Some comments arise from this discussion:

- To perform LS fits and estimate w_l^m in eq. (2.7), it is necessary to know coil sensitivities, so a calibration process is necessary as it is with SENSE.

- A key assumption is that the receiving coils are placed along the y direction, so eq. (2.7) may be solved for w_l^m: the main problem of SMASH is that it is not valid for general coil geometries.
- Although the reduction in Fourier average and the g–factor are not so immediate to identify as in SENSE, they both appear in this (and any other) reconstruction method.

Generalized SMASH (g–SMASH)

Generalized SMASH is based on SMASH, and it was intended to reconstruct images with arbitrary coil configurations and sub–sampling schemes [Byd02]. In this case, the reconstruction is done exclusively in the **k**–space. Similar to eq. (2.7), in g–SMASH the following relation is used:

$$W_l(\mathbf{x}) \simeq \sum_{m=-p}^{q} a_l^m \exp\left(j2\pi m\cdot\Delta k_y\cdot y\right), \tag{2.9}$$

for some a_l^m to be determined. If in eq. (2.4) the Fourier integral is performed only in the frequency encoding direction x, the signal in the "hybrid space" is obtained:

$$S(x,k_y) = \int_Q W_l(x,y)\rho(x,y)\exp(j2\pi k_y y)dy, \tag{2.10}$$

where the first variable of $S(x,k_x)$ is in the space domain and its second variable in the Fourier domain, hence the name "hybrid space". Once again, casting eq. (2.9) into eq. (2.10), the reconstructed signal reads:

$$\begin{aligned}\hat{S}(x,k_y) &= \int_{\mathrm{FOV}}\left(\sum_{m=-p}^{q} a_l^m \exp\left(j2\pi m\cdot\Delta k_y\cdot y\right)\right)\rho(x,y)\exp(j2\pi k_y y)dy\\ &= \sum_{m=-p}^{q} a_l^m \int_{\mathrm{FOV}}\rho(x,y)\exp\left(j2\pi\left(k_y+m\Delta k_y\right)y\right)\\ &= \sum_{m=-p}^{q} a_l^m S(x,k_y+m\Delta k_y) \simeq S(x,k_y).\end{aligned} \tag{2.11}$$

Without entering into unnecessary details, eq. (2.11) represents a system of coupled linear equations for each value of x and receiving coil l, which has to be solved to reconstruct the whole **k**–space. Although the corresponding matrix is highly sparse and therefore may be very efficiently inverted, for more complex sub–sampling schemes than keeping one of each r phase encoding lines its resolution may be computationally very intensive. This same problem has been reported for generalized SENSE, where non–uniform sub–samplings are considered [Pru01].

GeneRalized Auto–calibrating Partially Parallel Acquisition (GRAPPA)

GRAPPA is based on SMASH [Gri02], basing the reconstruction on the direct linear interpolation of the **k**–space **at each coil**, following:

$$\hat{s}_l(k_x,k_y+m\Delta k_y) \simeq \sum_j\sum_p \omega_{lj}(m-p)s_j(k_x,k_y+rp\Delta k_y), \tag{2.12}$$

for r the speedup, and p an index which select the sampled line to use. Lines at distances $rp\Delta k_y$ are actually acquired, while lines at distances $m\Delta k_y$ are not. The weights $\omega_{lj}(m-p)$ account for the distance between the line to estimate ($m\Delta k_y$) and the acquired line ($rp\Delta k_y$) through the dependence in $(m-p)^3$, and for actual spatial sensitivities of each coil through the double subindex lj.

Like in SMASH, these coefficients are computed with LS, but, instead of determining the weights to apply for the coil sensitivities to produce the line shifts on the **k**–space, GRAPPA aims to directly predict the value of missing lines in terms of the lines which have been measured. To that end, a set of lines at distance Δk_y are acquired, the so–called Auto Calibrating Signal (ACS), corresponding to the lowest frequencies of the specter in the direction of k_y. The weights ω_{lj} in eq. (2.12) are those which best fit the model, in the LS sense, for the calibrating lines. Theoretically, only one ACS line would be necessary (for $r=2$) to fit the model, but more lines are often acquired to achieve a better estimate. Besides, since those lines in the center of the specter are used, the SNR is the highest possible (since the most of the energy of the image is concentrated in these frequencies), so the estimate is quite robust. Furthermore, ACS lines are part of the acquired signal (in fact, they are used to compute the DFT and obtain the **x**–space), so the algorithm is auto–calibrated.

As opposed to the techniques previously introduced, GRAPPA shows a characteristic which clearly distinguishes it: eq. (2.12) holds for each receiving coil, so instead of a single **k**–space GRAPPA provides one **k**–space per antenna, and thus one **x**–space (one image) per receiving coil. This feature assimilates GRAPPA reconstructed signals to systems with multiple coils without **k**–space subsampling. As a result, the multiple images have to be combined to obtain the final composite image. Since the signal in the **k**–space is arbitrary, the image in the **x**–space is in general complex, so in practice the information of interest is the modulus of this signal. With these two considerations, it has been shown in [Gil07] that the optimal combination in the sense of LS, when coil sensitivities cannot be characterized, is given by Sum of Squares (SoS):

$$A_L(\mathbf{x}) = \sqrt{\sum_{l=1}^{L} |S_l(\mathbf{x})|^2}, \qquad (2.13)$$

for $S_l(\mathbf{x})$ the complex signal in the **x**–space for each coil, and $A_L(\mathbf{x})$ the real–valued Composite Magnitude Signal (CMS). For single coil systems, eq. (2.13) is still valid when particularized to the case $L=1$. Obviously, in this case, this expression yields simply the modulus of the complex signal in the **x**–space. Although in this case the signal is not composite strictly speaking, the general denomination of CMS will be employed throughout this dissertation to refer to the modulus of the signal in the image domain, proceeding either from one single coil or multiple coils.

As a final remark, GRAPPA has two characteristics that make it a very attractive approach compared to the other methods presented:

- It does not require to characterize coil sensitivities, so no additional calibration is needed (the algorithm is auto-calibrated).
- It does not require either to invert any linear system of equations, and is equally simple for arbitrary coil configurations and sub–sampling schemes.
- Non–uniform subsampling schemes are equally simple to use with the interpolation described by eq. (2.12), as opposed to the other pMRI techniques.

Although SENSE may be preferable if coil sensitivities may be characterized [Lar07], GRAPPA is fast and simple, which probably is the reason why commercial parallel MRI scanners tend to integrate GRAPPA for image reconstuction.

---

[3] Although in general the weights are $\omega_{lj}(m,p)$, for practical applications a stationary behavior is assumed, so the weights reduce to $\omega_{lj}(m-p)$.

2.5 Diffusion Weighted Imaging

In what follows, the discussion is centered on Diffusion MRI. This modality goes beyond anatomical characterization, being mainly useful (but not limited to) describe the connectivity of nervous fibers in the white matter of the brain. As shown in Fig. 2.1, conventional MRI allows the study of soft tissues of the brain, allowing to distinguish gray matter from white matter. The white matter is highly structured, comprising the nervous fibers (neural axons) which connect different parts of the brain. The imaging of these structures is carried out taking advantage of the diffusion of water molecules. Without restrictions, water molecules follow a random movement known as Brownian motion. If a particle is at position $\mathbf{x}$ at time t, at time $t+\Delta t$ Brownian motion pushes it to a random position $\mathbf{x}'$, which is Gaussian distributed with mean $\mathbf{x}$ and isotropic variance proportional to Δt. In the white matter, water molecules are confined within neural axons, which are surrounded by myelin coats avoiding water diffusion trough them. Hence, the diffusion process is not isotropic, but instead it is preferentially produced along the direction of the nervous fiber, in which no myelin barriers are found. This is the principle used by DWI to infer **macroscopic** fiber architectures.

Despite these considerations, the Gaussian model cannot be assumed inside myelin coats. The statistics of restricted diffusion, even for very simple scenarios, are much more complicated. In [Söd95], the neural axon is modeled as a finite bounded cylinder; the diffusion inside the cylinder is Gaussian, but its walls avoid any diffusion outsides. This model has been recently used in some works on Diffusion Imaging [Öza06, Prč08]. However, from eq. (2.4), it is easy to note that MRI scanners do not model the behavior of each water molecule, but the macroscopic behavior, or ensemble average, of all the spins within the voxel resolution, described by the spin density $\rho(\mathbf{x})$. The probability density of the ensemble average, commonly known as ensemble–average diffusion propagator, is:

$$P(\mathbf{x}'|t) = \int_{\Omega} \rho(\mathbf{x}) P_s(\mathbf{x}'|\mathbf{x},t) d\mathbf{x}, \tag{2.14}$$

where $P_s(\mathbf{x}'|\mathbf{x},t)$ is the conditional Probability Density Function (PDF) for each spin, the so–called self–diffusion probability, which may be modeled as in [Söd95], and Ω is the physical space of a voxel. It is clear from eq. (2.14) that $P_s(\mathbf{x}'|\mathbf{x},t)$ does not represent the macroscopic PDF of the diffusion process. The PDF at the voxel level, $P(\mathbf{x}'|t)$, may be seen as the marginalization in $\mathbf{x}$ of a mixture of independent and (roughly) identically distributed individual spin PDF. Typical voxel sizes are in the order of millimeters, while axon diameters are in the order of microns. Inside each axon, a high number of particles will be present, so it is easy to notice that the mixture of statistics comprises thousands of water molecules, each of them following a similar PDF. The central limit theorem allows to conclude that a Gaussian model accurately describes macroscopic diffusion whenever all fiber bundles in the space of the voxel follow similar statistics, i.e. they are oriented in the same direction. This assumption will be discussed in Section 2.7.

2.5.1 The Pulse Gradient Spin Echo experiment

The most important conclusion from the previous paragraphs is that Diffusion Imaging measures the macroscopic diffusion of water molecules. The confinement of particles in myelin coats creates preferred directions for diffusion, which are associated to the directions of fiber bundles. In other words, the probability of diffusion in a given direction inside a voxel may be identified with the probability of the existence of a fiber bundle in this same direction. Given the macroscopic

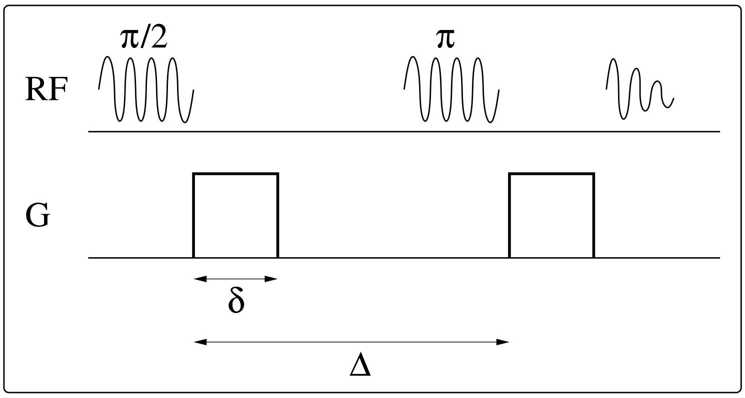


Figure 2.5: Magnetic field gradients for the Pulse Gradient Spin Echo experiment. After the RF pulse $\mathbf{B}_1(t)$ is applied, the first pulsed diffusion gradient is activated. Like G_y for phase encoding of the **k**–space, this pulse induces a phase–shift in the precession of the spins. The additional RF pulse inverts the phase of the precession before a second pulsed diffusion gradient is applied. The molecules which have moved along the diffusion direction **g** are affected by a different phase–shift for the first and the second pulse, so a net phase–shift is induced for them. For the particles which cannot move along **g**, the three phase–shifts compensate and the received signal is in phase with the baseline T_2 image without diffusion gradients.

interpretation of the diffusion process, the analysis of nervous fibers has to be seen in a macroscopic sense as well: single fibers cannot be characterized with DWI, but only groups of fibers with similar diffusion statistics.

For this reason, it is interesting to relate the ensemble–average diffusion propagator with the signal measured by the MRI scanner. The basic physical principle in this case was firstly investigated by Hanh [Han50]: in the presence of a magnetic field inhomogeneity, thermal motion of the spins lead to the attenuation of spin echoes. This work was used by Stejskal and Tanner [Ste65] to explicitly measure the diffusion in the Pulse Gradient Spin Echo (PGSE) experiment, schematically described in Fig. 2.5. From eq. (2.1), the Larmor frequency is proportional to the strength of the magnetic field, so, when the first pulsed gradient is applied, a spatial–dependent deviation in the frequency of spin precession is induced:

$$\omega_0'(\mathbf{x}) = \gamma\left(\mathbf{B}_0 + \mathbf{G}^T\mathbf{x}\right) = \omega_0 + \gamma\mathbf{G}^T\mathbf{x}, \tag{2.15}$$

where the total deviation of the magnetic field with respect to $\mathbf{B}_0$ is its directional derivative in the desired direction, i.e. the projection of the gradient $\mathbf{G}$ on $\mathbf{x}$, $\mathbf{G}^T\mathbf{x}$. After a time δ, the pulsed gradient disappears, so the rotation frequency for all spins is again ω_0. Withal, since the spins have been rotating with different velocities, their phases are different:

$$\phi(\mathbf{x}) = \int_0^{\delta} \omega_0'(\mathbf{x})dt = \delta\omega_0 + \gamma\delta\mathbf{G}^T\mathbf{x}, \tag{2.16}$$

and the phase difference with respect to the T_2 image[4], for which no gradient is applied, is simply:

$$\varphi(\mathbf{x}) = \gamma\delta\mathbf{G}^T\mathbf{x}. \tag{2.17}$$

The 180^o RF pulse is used to invert the phase of the spins:

$$\varphi'(\mathbf{x}) = -\gamma\delta\mathbf{G}^T\mathbf{x}. \tag{2.18}$$

The second pulsed gradient is applied a time Δ after the first one. It is assumed that $\Delta \gg \delta$, so the motion of particles during the application of the pulses is negligible compared to the motion during the time between pulses (and the motion

---

[4] Note that the PGSE experiment is only feasible for the T_2 modality, since it is based on phase–shifts of the transverse component.

during δ is negligible compared to $\mathbf{G}^T\mathbf{x}$). This is known as the **narrow pulses condition**. After the time $\varDelta$, the particle moves to a random position $\mathbf{x}'$ described by the conditional (self–diffusion) probability $P_s(\mathbf{x}'|\mathbf{x},t)$. Consequently, it is affected by a different deviation in the Larmor frequency and hence a different phase deviation than it was as a consequence of the first pulse. From eq. (2.17), and taking into account that the particle is now at position $\mathbf{x}'$, with an additional phase deviation of 180^o (due to the radio frequency pulse), the total phase deviation is:

$$\varphi''(\mathbf{x}) = \gamma\delta\mathbf{G}^T\mathbf{x}' - \gamma\delta\mathbf{G}^T\mathbf{x} = \gamma\delta\mathbf{G}^T(\mathbf{x}' - \mathbf{x}), \tag{2.19}$$

with respect to the measured signal for which no gradients are applied. If the gradient is applied in a transverse direction to the fiber bundle, then $\mathbf{G}^T(\mathbf{x}' - \mathbf{x}) = 0$, since the particle has to move along fiber directions, and no phase–shift is induced. The RF signal is acquired after the second pulse disappears, and given the result in eq. (2.19), it will be related to the original T_2 signal (when no gradients are applied) as:

$$S(\mathbf{G}) = S(\mathbf{0})\exp\left(j\gamma\delta\mathbf{G}^T(\mathbf{x}' - \mathbf{x})\right) \Leftrightarrow S(\mathbf{G})/S(\mathbf{0}) = \exp\left(j\gamma\delta\mathbf{G}^T(\mathbf{x}' - \mathbf{x})\right), \tag{2.20}$$

where S is the acquired signal. Obviously, eq. (2.20) holds for each particle, but given the macroscopic nature of DWI the measured signal will be the superposition of phase–shifted spins for all the molecules at voxel resolution. Thence, only the expected value of eq. (2.20) can be computed:

$$S(\mathbf{G})/S(\mathbf{0}) = E\left\{\exp\left(j\gamma\delta\mathbf{G}^T(\mathbf{x}' - \mathbf{x})\right)\right\}. \tag{2.21}$$

2.5.2 The q–space

The expectation in eq. (2.21) may be computed in terms of the diffusion propagator. Indeed, the expectation has to be calculated for the joint PDF in both $\mathbf{x}$ and $\mathbf{x}'$ to account for the macroscopic effect inside the voxel, for all possible initial positions $\mathbf{x}$, and final positions $\mathbf{x}'$, after a time $\varDelta$:

$$S(\mathbf{G})/S(\mathbf{0}) = \iint P(\mathbf{x},\mathbf{x}'|\varDelta)\exp\left(j\gamma\delta\mathbf{G}^T(\mathbf{x}' - \mathbf{x})\right)d\mathbf{x}d\mathbf{x}'. \tag{2.22}$$

In Diffusion Imaging, it is not so much the final position $\mathbf{x}'$ as the displacement $\mathbf{R} = \mathbf{x}' - \mathbf{x}$ what results of interest. In fact, it is the probability of a displacement $\mathbf{R}$ in a given direction $\mathbf{r} = \|\mathbf{R}\|$ which may be related to the existence of fiber bundles along $\mathbf{r}$, and not the absolute position $\mathbf{x}'$. On the other hand, the joint probability may be computed as in eq. (2.14), which is a marginalization of this joint PDF, so eq. (2.22) may be written:

$$S(\mathbf{G})/S(\mathbf{0}) = \iint \rho(\mathbf{x})P_s(\mathbf{x}+\mathbf{R}|\mathbf{x},\varDelta)d\mathbf{x}\exp\left(j\gamma\delta\mathbf{G}^T\mathbf{R}\right)d\mathbf{R}. \tag{2.23}$$

At this point, it is useful to introduce the transformed variable $\mathbf{q}$, much like it is done with the variable $\mathbf{k}$:

$$\mathbf{q} = \frac{1}{2\pi}\gamma\delta\mathbf{G}, \tag{2.24}$$

which is used to define the so–called $\mathbf{q}$–space as the (scaled) space of all possible gradient directions and magnitudes for the PGSE experiment. On the other hand, the marginalization used in eq. (2.14) to obtain the ensemble–average diffusion propagator may be applied to eq. (2.23):

$$\begin{aligned} S(\mathbf{G})/S(\mathbf{0}) &= \int \exp\left(j2\pi\mathbf{q}^T\mathbf{R}\right)\left(\int \rho(\mathbf{x})P_s(\mathbf{x}+\mathbf{R}|\mathbf{x},\varDelta)d\mathbf{x}\right)d\mathbf{R} \\ &= \int \exp\left(j2\pi\mathbf{q}^T\mathbf{R}\right)P(\mathbf{R}|\varDelta)d\mathbf{R} = \mathfrak{F}^{-1}\left\{P(\mathbf{R}|\varDelta)\right\}(\mathbf{q}), \end{aligned} \tag{2.25}$$

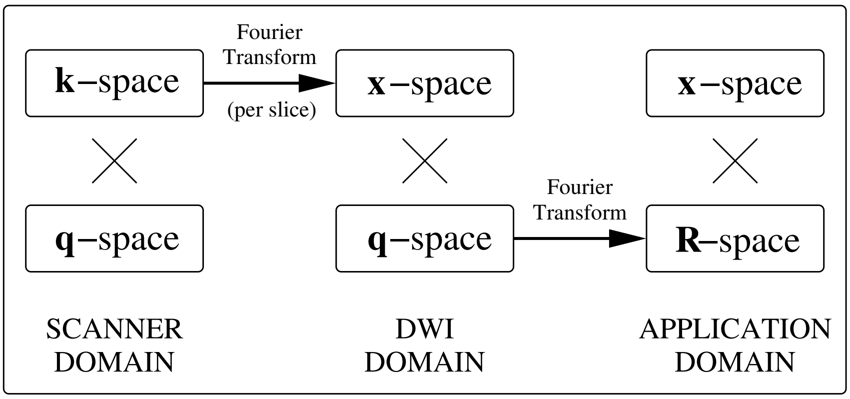


Figure 2.6: Diffusion Imaging as a 6D process. The acquisition is performed by the scanner in the dual **k**–space for each pulsed gradient **G**, which univocally determines **q**, so the scanner domain is the Cartesian product of these two spaces. The **x**–space (image domain) is obtained, for the frequency/phase encoding, as the 2D Fourier transform of each slice z_p of the **k**–space. This transformation may be accomplished for each value of the **q**–space, so the resulting signal in the DWI domain is the Cartesian product of these two spaces (obviating the computation of the modulus to obtain the CMS). Alternatively, a whole **q**–space is defined for each location **x**, so the diffusion propagator in the **R**–space is defined in turn for each **x**.

where $\mathfrak{F}^{-1}\{f(\mathbf{R})\}(\mathbf{q})$ represents the inverse Fourier transform of $f(\mathbf{R})$ in the dual variable $\mathbf{q}$. In fact, eq. (2.25) represents the well–known Fourier relationship between measured signals and displacement probabilities under the short pulses condition [Cal91]. In practice, it can be demonstrated that it is necessary to characterize only the modulus of $S(\mathbf{G})/S(\mathbf{0})$, so its phase can be neglected [Wed05]. Accordingly, the signal of interest is given by:

$$E(\mathbf{x},\mathbf{q}) = \left|\frac{S(\mathbf{x},\mathbf{G})}{S(\mathbf{x},\mathbf{0})}\right| = \frac{|S(\mathbf{x},\mathbf{G})|}{|S(\mathbf{x},\mathbf{0})|}, \tag{2.26}$$

where $|S(\mathbf{x},\mathbf{G})|$ and $|S(\mathbf{x},\mathbf{0})|$ are obviously the corresponding CMS at $\mathbf{x}$ when a gradient $\mathbf{G}$, or no gradient at all, are applied. In the previous equation, the spatial dependency has been included to stress the fact that eq. (2.25) describes the diffusion process at each voxel. Finally, eq. (2.25) is equivalent to:

$$P(\mathbf{R}|\Delta) = \mathfrak{F}\{|E(\mathbf{q})|\}(\mathbf{R}), \tag{2.27}$$

which is the basis for Diffusion MRI: the PDF of the displacement of water molecules in the extent of a voxel can be computed as the 3D Fourier transform of $E(\mathbf{q})$, which is the CMS provided by the MRI scanner at each image location and for each sensitizing gradient $\mathbf{G} = 2\pi/\gamma\delta \cdot \mathbf{q}$. The data set comprising, at each image location $\mathbf{x}$, a whole set of measurements of the CMS, $|S(\mathbf{G})|$, for a number of independent sensitizing gradients $\mathbf{G}$, is commonly referred to as Diffusion Weighted Image (DWI).

2.5.3 Diffusion Imaging as a 6D imaging technique

The acquisition of the **q**–space is independent on the acquisition of the **k**–space. The pulse sequence depicted in Fig. 2.5 has to be combined with the pulse sequence used to code spatial information, so the whole **k**–space is acquired with each of the diffusion gradients **G**. Alternatively, each point in the **k**–space has to be acquired for all possible values of **G**. Given the relation between the **k**–space and the image domain (the **x**–space), it is obvious that, for each spatial location **x**, all pulsed gradient sequences **G** are defined as well, see eq. (2.26). From the relationship between acquired signals in the **q**–space and diffusion propagators, it may be concluded that the 3D diffusion propagator is defined for each 3D spatial location $\mathbf{x}$: at each image voxel $\mathbf{x}$, the diffusion propagator $P(\mathbf{R}|\Delta)$ gives the probability of diffusion across this voxel to a point $\mathbf{x}+\mathbf{R}$. All these relations are summarized in Fig. 2.6.

Diffusion Imaging may be seen as the 6D Cartesian product between the 3D $\mathbf{x}$–space and the 3D $\mathbf{R}$–space of ensemble–average diffusion propagators. At the same time, the dual Fourier spaces on variables $\mathbf{k}$ and $\mathbf{q}$ form another 6D space. Since the acquisition of the $\mathbf{k}$–space and the $\mathbf{q}$–space respond to different physical processes, Fourier analysis may be used to obtain the Cartesian product between the $\mathbf{x}$–space and the $\mathbf{q}$–space, which is the domain of DWI. Besides, any existing technique to acquire the $\mathbf{k}$–space, including pMRI, may be used in Diffusion Imaging. Corresponding to the work–flow in Fig. 2.6, the following problems may be identified:

1. Acquisition of the $\mathbf{k}\times\mathbf{q}$–space. The design of pulse sequences and transmitting/receiving coils is beyond the scope of this dissertation. On the other hand, the accelerated acquisition of the $\mathbf{k}$–space from multiple–coil systems, although still an open problem, has been partially solved with pMRI.

2. The reconstruction of the $\mathbf{x}$–space, when frequency/phase encoding is used, may be easily performed with DFT techniques. Therefore, the construction of DWI signals is not an issue for the work presented here.

3. The reconstruction of the $\mathbf{R}$–space from the $\mathbf{q}$–space is an open problem, since it is computationally prohibitive to sample the whole $\mathbf{q}$–space and use the DFT on sampled data. In Section 2.6, Diffusion Tensor Imaging (DTI) is presented as an analytical solution whenever a Gaussian model may be assumed. In Section 2.7, it is stated that the Gaussian model is very limited in a variety of scenarios. Accordingly, in a general situation, the reconstruction of the $\mathbf{R}$–space from a sparse sampling of the $\mathbf{q}$–space has not a closed solution. This is the main focus of this dissertation, which will be addressed in Chapters 6 and 7.

4. Finally, all this signal–processing tasks are embedded in a noisy environment. The propagation and effects of noise on each stage depend on a number of factors, such as the reconstruction scheme for the $\mathbf{k}$–space and the $\mathbf{R}$–space. These problems are as well addressed in this dissertation. In Chapter 3, the properties of noise in the $\mathbf{k}$ and $\mathbf{x}$–spaces are analyzed. Chapter 4 is devoted to the analysis of the propagation of noise from the $\mathbf{q}$–space to the $\mathbf{R}$–space. In Chapter 5, a number of techniques in the $\mathbf{x}\times\mathbf{q}$–space (DWI domain) are introduced to overcome these adverse effects.

2.6 Diffusion Tensor Imaging

The inference of exact information on the $\mathbf{R}$–space would require the sampling of the whole $\mathbf{q}$–space to use the Fourier relationship in eq. (2.27). Even so, the assumption of a Gaussian diffusion propagator allows to analytically solve the Fourier integral, and thus to easily formalize water diffusion from a few DWI data sets [Bas94a,Bas96]. From eq. (2.14), it may be inferred that a Gaussian diffusion propagator can be assumed in case all fiber bundles within the voxel resolution are nearly identically distributed. In this case, $P(\mathbf{R}|\Delta)$ is a mixture of independent and (nearly) identically distributed bounded cylinder statistics and, by virtue of the central limit theorem, their superposition is Gaussian distributed:

$$P(\mathbf{R}|\Delta) = \frac{1}{\sqrt{|\mathscr{D}|(4\pi\Delta)^3}} \exp\left(\frac{-\mathbf{R}^T\mathscr{D}^{-1}\mathbf{R}}{4\Delta}\right). \tag{2.28}$$

From eq. (2.25), the measured signal in the **q**–space is the Fourier transform of the PDF in eq. (2.28):

$$E(\mathbf{q}) = \mathfrak{F}^{-1}\left\{P(\mathbf{R}|\Delta)\right\}(\mathbf{q}) = \exp\left(-4\pi\tau\mathbf{q}^T\mathscr{D}\mathbf{q}\right), \tag{2.29}$$

which is the well–known Stejskal–Tanner equation [Ste65]. In the previous equation, it is not necessary to compute the modulus of $E(\mathbf{q})$, since it is already real and strictly positive. The effective diffusion time $\tau = \Delta - \delta/3$ includes a correction $\delta/3$ in the diffusion time Δ to compensate the diffusion during the appliance of the pulse. The diffusion tensor $\mathscr{D}$ is the anisotropic covariance matrix of the Gaussian PDF $P(\mathbf{R}|\Delta)$, and therefore is a symmetric, positive–definite matrix, with positive eigenvalues and orthonormal eigenvectors. Eq. (2.29) is usually re–written in terms of the b–value, which is a typical parameter in DTI:

$$\begin{aligned} E(\mathbf{g}) &= \exp\left(-b\mathbf{g}^T\mathscr{D}\mathbf{g}\right); & (2.30)\\ b &= 4\pi^2\tau\|\mathbf{q}\|^2 = \tau\gamma^2\delta^2\|\mathbf{G}\|^2/\pi. & (2.31)\end{aligned}$$

The interest of eq. (2.30) is that E does not depend on $\mathbf{q}$, but only on its direction $\mathbf{g} = \mathbf{G}/\|\mathbf{G}\| = \mathbf{q}/\|\mathbf{q}\|$ (note that the b–value depends only on the scanning parameters: the duration and spacing between the pulses, the strength of the diffusion gradients, and the gyromagnetic ratio). At a given voxel, the diffusion may be characterized in terms of only six parameters, the six free components of the diffusion tensor. As a consequence, once the signal in the **q**–space has been sampled for a reduced number of gradient directions (a minimum of six), a linearized version of eq. (2.30) may be used to solve for the components of the diffusion tensor:

$$-\frac{\log E(\mathbf{g})}{b} = g_1^2\mathscr{D}_{11} + 2g_1g_2\mathscr{D}_{12} + 2g_1g_3\mathscr{D}_{13} + g_2^2\mathscr{D}_{22} + 2g_2g_3\mathscr{D}_{23} + g_3^2\mathscr{D}_{33}, \tag{2.32}$$

for $\mathbf{g} = [g_1, g_2, g_3]^T$, which defines a system of linear equations if a set of independent gradient directions $\{\mathbf{g}_i\}_{i=1}^N$ is used. If more than $N = 6$ gradients are available, the system is overdetermined, and is often solved with LS techniques [Sal05].

2.6.1 Interpretation of tensor data

DTI provides a complete characterization of the diffusion process whenever the Gaussian model holds. However, the interpretation of tensor data is not so trivial as in the case of anatomical MRI. In this Section, the techniques for visualization, interpretation, and quantitative measurements for DTI are briefly reviewed.

Scalar parameters

A number of significant scalar measures may be extracted from tensor data, usually from eigenvalues and eigenvectors representations. In what follows, the three eigenvalues of the tensor are assumed to be $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$, associated to orthonormal eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$. The following groups of measurements may be distinguished [Wes01]:

Amount of diffusion. The Mean Diffusivity (MD) is defined as the mean value of the diffusion along each principal direction $\mathbf{e}_i$, and gives an idea of how easily

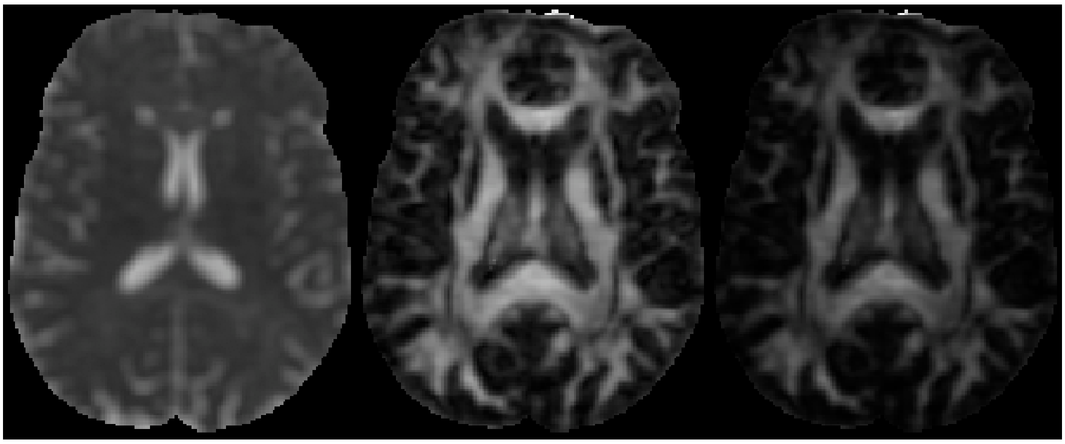


Figure 2.7: Scalar parameters in DTI: mean diffusivity (left), fractional anisotropy (center), and relative anisotropy (right) for an axial slice of data set CDR1.

water molecules move in a given tissue:

$$\mathrm{MD}(\mathscr{D}) = \frac{\lambda_1+\lambda_2+\lambda_3}{3}. \tag{2.33}$$

Anisotropy measures. These indices measure the difference in diffusion capability between the preferred direction $\mathbf{e}_1$ and the remaining directions. A high anisotropy corresponds to regions where fiber bundles are highly structured in well defined directions (corresponding to $\mathbf{e}_1$). Low anisotropies correspond to regions corresponding either to unstructured tissue or to fiber bundles crossing in multiple directions. Two indices may be defined; the Fractional Anisotropy (FA):

$$\mathrm{FA}(\mathscr{D}) = \frac{1}{\sqrt{2}}\sqrt{\frac{(\lambda_1-\lambda_2)^2+(\lambda_1-\lambda_3)^2+(\lambda_2-\lambda_3)^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2}}, \tag{2.34}$$

and the Relative Anisotropy (RA):

$$\mathrm{RA}(\mathscr{D}) = \frac{1}{\sqrt{2}}\frac{\sqrt{(\lambda_1-\lambda_2)^2+(\lambda_1-\lambda_3)^2+(\lambda_2-\lambda_3)^2}}{\lambda_1+\lambda_2+\lambda_3}. \tag{2.35}$$

Both indices lay in the range between 0 (isotropic diffusion) and 1 (completely anisotropic diffusion, with some $\lambda_i = 0$). Fig. 2.7 shows an example of the representations obtained with these parameters for a real data set.

Shape Coefficients. Although anisotropy measures describe the directional behavior of diffusion, they are prone to ambiguities in some cases. For example, for the combinations: $\lambda_1 = 1,\ \lambda_2 = 1,\ \lambda_3 = 0$ and $\lambda_1 = 1,\ \lambda_2 = 1/4,\ \lambda_3 = 1/4$ the FA is in both cases $1/\sqrt{2}$, but they correspond to completely different situations. In the former, the diffusion is isotropic in the plane formed by $\mathbf{e}_1$ and $\mathbf{e}_2$, while no diffusion is allowed in the direction $\mathbf{e}_3$. The diffusion in this case in ‘planar’, since it proceeds in a plane. In the latter, diffusion does preferentially along the direction $\mathbf{e}_1$, corresponding to ‘linear’ diffusion. To resolve this ambiguity, the following shape coefficients may be defined [Wes01]:

$$c_l = \frac{\lambda_1-\lambda_2}{\lambda_1};\ c_p = \frac{\lambda_2-\lambda_3}{\lambda_1};\ c_s = \frac{\lambda_3}{\lambda_1}, \tag{2.36}$$

where c_l is the linear, c_p the planar, and c_s the spherical coefficient. Fig. 2.8 shows an example of the representations obtained with these parameters.

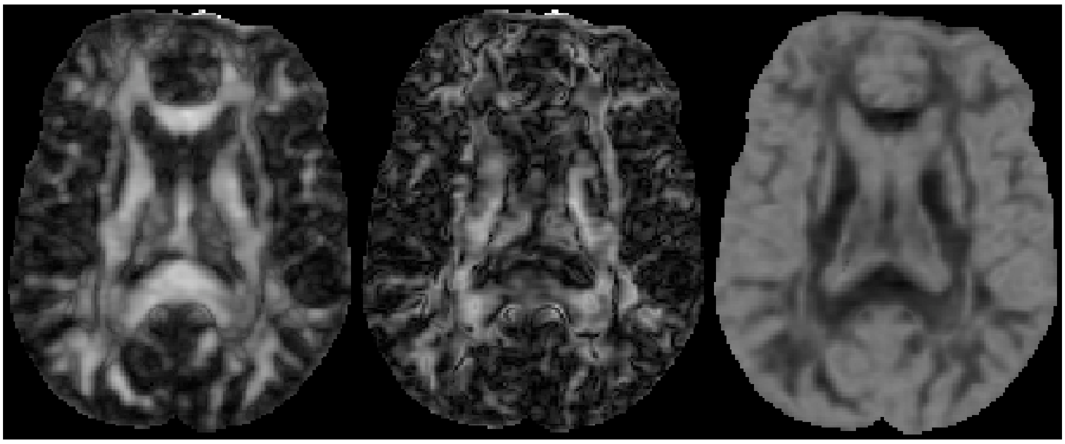


Figure 2.8: Scalar parameters in DTI: linear (left), planar (center), and spherical (right) coefficients for an axial slice of data set CDR1.

Color coding

The aim of color coding is to represent more reach information than with scalar coefficients. To that end, it is assumed that most of the voxels of the image (or the slice which is being represented) correspond to linear diffusion (high FA and c_l), and therefore the tensor may be accurately described by the direction of its principal eigenvector, $\mathbf{e}_1$. This is the same as to say that each voxel is only crossed by one group of fiber bundles in the same direction, which is approximately true in most of cases. Color coding techniques decompose $\mathbf{e}_1$ in three orthogonal components, corresponding to physical coordinates 'x' (sagittal), 'y' (coronal), and 'z' (axial). These three components are scaled depending on the FA for each voxel, and plotted as the RGB (respectively) channels of the colored image, as shown in Fig. 2.9.

Glyphs

Glyphs are 3D plots which aim to represent the whole diffusion process. Although many glyph drawings are possible [MC08], one very popular approach

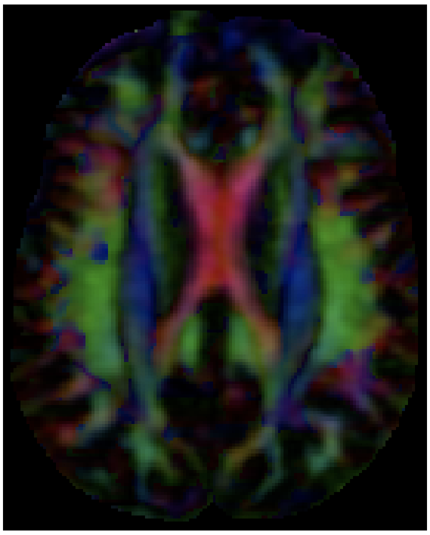


Figure 2.9: DTI representation: an example of color–coding over real data set CDR1.

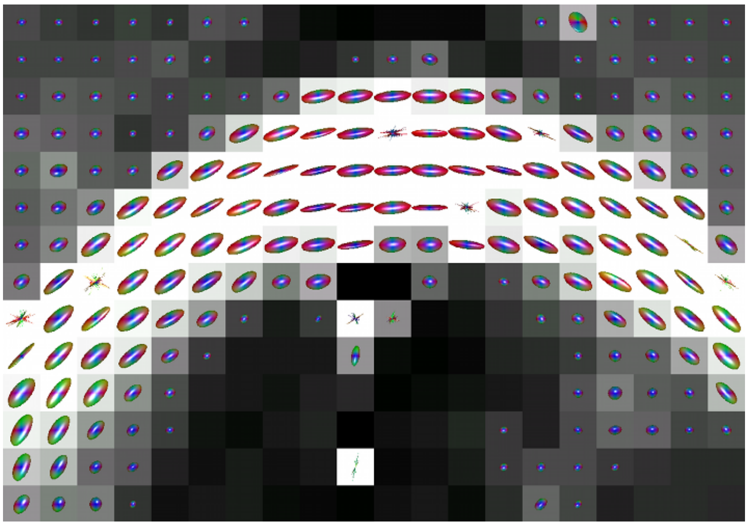


Figure 2.10: DTI representation: an example of glyphs, combined with an FA map, to represent the corpus callosum in a human brain (data set CDR1).

is the explicit representation of $P(\mathbf{R}|\Delta)$, or some orientation information associated to this PDF, in a spherical coordinates plot. For example, the radial integral of $P(\mathbf{R}|\Delta)$ along each direction $\mathbf{r} = \mathbf{R}/\|\mathbf{R}\|$, so–called Orientation Distribution Function (ODF) [Tuc03, Tuc04], may be computed to obtain the orientation information related to the 3D PDF. At each voxel, a surface is rendered with the following parametrization:

$$\begin{bmatrix} x(\theta,\phi) \\ y(\theta,\phi) \\ z(\theta,\phi) \end{bmatrix} = \begin{bmatrix} \Psi(\theta,\phi)\sin(\theta)\cos(\phi) \\ \Psi(\theta,\phi)\sin(\theta)\sin(\phi) \\ \Psi(\theta,\phi)\cos(\theta) \end{bmatrix}, \tag{2.37}$$

where $\Psi(\theta,\phi) \equiv \psi(\mathbf{r})$ is the ODF, and spherical coordinates follow physics convention. In other words, for each orientation $\mathbf{r}$, the amount of diffusion is represented as the distance from the surface to the origin (placed in the center of each image voxel). For DTI, this representation is an ellipsoid at each voxel, whose principal axis are aligned with the eigenvectors $\mathbf{e}_i$, and their lengths correspond to the eigenvalues λ_i [Tuc04]. As a final remark, glyphs may be combined with color coding, anisotropy maps, or any other information to give richer information. Fig. 2.10 shows an example of this kind of representation.

Tractography

The most widely used application of DTI is tractography or fiber tracking. It consists in the estimation of the trajectories of fiber bundles from diffusion tensors. In practice, a number of seeding points are placed at selected voxels, and for those of them for which a certain criterion is met (usually if their FA or linear coefficient c_l is over a certain threshold), the following point in the path is found, following the direction of the principal eigenvector $\mathbf{e}_1$. This new point is used in turn as a seeding point to follow the path, until the FA no longer meets the criterion. To regularize this process, fiber tracking algorithms are commonly based on robust numerical integration of the vector field formed by $\mathbf{e}_1$ [Bas00b, Ten02]. Since the diameter of nervous fibers is in the order of microns, and the voxel size in the order of millimeters, it remains evident that fiber tracking will not be able to follow single fibers. Fortunately, the white matter is structured in coarse fiber bundles whose macroscopic behavior is accurately described, in most of cases, by the diffusion tensor, and tractography is able to find good estimates of these fiber paths. Besides the rendering of fiber bundles, fiber tracking allows to trace connectivity

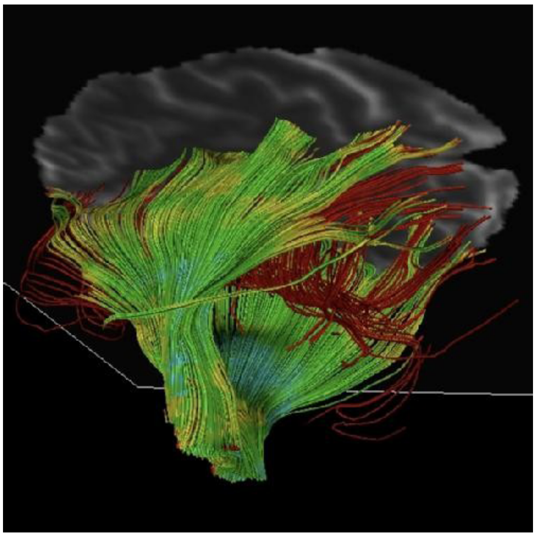


Figure 2.11: DTI representation: an example of fiber tracking using a Runge–Kutta method of order 4, over data set CDR1. The pyramidal tract and the corpus callosum have been traced.

maps between different regions [Fri06]. A representative example is depicted in Fig. 2.11.

2.7 Limitations of DTI. High Angular Resolution Diffusion Imaging

The fundamental hypothesis underlying DTI is the Gaussian behavior of $P(\mathbf{R}|\Delta)$. The ensemble–average described by eq. (2.14) allows to use the central limit theorem to infer the Gaussianity of $P(\mathbf{R}|\Delta)$ when the individual diffusion propagators may be considered as identically distributed. At the same time, if $P(\mathbf{R}|\Delta)$ is normally distributed, the well–known Stejskal–Tanner relation in eq. (2.30) holds, and the sampling of $E(\mathbf{q})$ for a unique b–value is enough to characterize the whole function: suppose that $E(\mathbf{q})$ has been sampled for $b = b_0$; then, for any other b–value:

$$E(\mathbf{g};b') = \exp\left(-b'\mathbf{g}^T\mathscr{D}\mathbf{g}\right) = \exp\left(-b\frac{b'}{b}\mathbf{g}^T\mathscr{D}\mathbf{g}\right) = E(\mathbf{g};b)^{b'/b}, \tag{2.38}$$

and the whole $E(\mathbf{q})$ may be computed from one sampled sphere with radius q_0; the diffusion propagator may then be computed from eq. (2.27) to yield eq. (2.28).

Still, there is a number of situations where the Gaussian model cannot be assumed. The most classical one occurs when two different, well–defined fiber bundles cross in different directions. It is obvious that in this situation the individual diffusion propagators P_s cannot be considered identically distributed, since they are clustered around two dominant directions. Two modes appear in the ensemble average diffusion propagator, which no longer fits the Gaussian model. This is one possible origin of low FA tensors. If, even so, the tensor model is used to fit data, two similar eigenvalues $\lambda_1 \simeq \lambda_2$ appear. It translates in the ambiguity in the orientation of $\mathbf{e}_1$ and $\mathbf{e}_2$, so it is impossible to infer the actual orientation of the fiber bundles. Fig. 2.12 illustrates this behavior. When the tensor model does not hold, the signal in the $\mathbf{q}$–space may always be written:

$$E(\mathbf{q}) = \exp\left(-bD(b,\mathbf{g})\right), \tag{2.39}$$

where D is the Apparent Diffusion Coefficient (ADC). Comparing eqs. (2.38) and

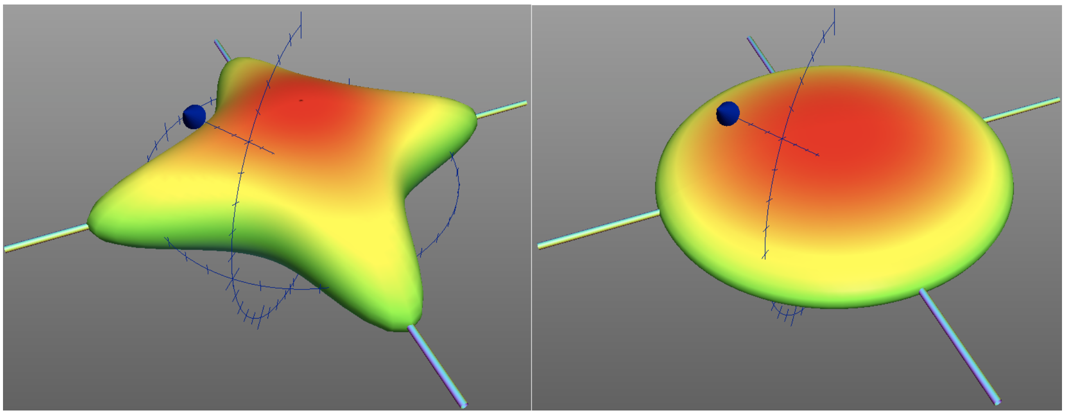


Figure 2.12: An example of fiber architecture which cannot be modeled with DTI. Two fiber bundles, in orthogonal directions and with the same partial volume fractions, cross in the same voxel. On the left–hand side, the true ODF is represented, which correctly accounts for the two modes in the distribution. On the right–hand side, a tensor model has been fitted originating to equal eigenvalues $\lambda_1 = \lambda_2 > \lambda_3$. There is an ambiguity in the orientation of $\mathbf{e}_1$ and $\mathbf{e}_2$ which makes it impossible to infer the actual directions of fiber bundles (depicted in the figure) from the resulting ODF.

(2.39), the positive–definite quadratic form $\mathbf{g}^T\mathscr{D}\mathbf{g}$ is replaced by the positive function $D(\mathbf{g},b)$ which, in this case, depends on the modulus of $\mathbf{q}$ through b. This implies that the reasoning in eq. (2.38) does not hold for non–Gaussian diffusion propagators, and therefore the signal in the **q**–space cannot be completely characterized from data sampled with a unique b for the general case. Explicit Fourier analysis has to be performed in this case to infer the behavior of $P(\mathbf{R}|\Delta)$.

A number of techniques have recently appeared to deal with this problem. The most obvious one is to sample the whole **q**–space (i.e. for all possible orientation and modulus of **q**), and numerically estimate $P(\mathbf{R}|\Delta)$ with DFT techniques. This approach is called Diffusion Spectrum Imaging (DSI) [Tuc03, Tuc04, Wed05]. The main drawback of DSI is the need to sample the whole **q**–space, since it may take prohibitive times for clinical practice. Instead, it is common to use a suboptimal sampling scheme known as High Angular Resolution Diffusion Imaging (HARDI), for which the **q**–space is only sampled for a unique b–value, but for a large number of gradient directions **g**. The minimum number of gradient directions for DTI is 6; with HARDI techniques, it is common to acquire 50, 100 or even 200 directions. Although the complete characterization of $P(\mathbf{R}|\Delta)$ is impossible, HARDI techniques aim to describe only a certain orientation information like the ODF, not through the calculation of $P(\mathbf{R}|\Delta)$ and the subsequent integration in $\mathbf{R}$, but by means of some indirect method. This method has to deal with the incomplete information provided by HARDI sampling, either including some prior knowledge, or making certain assumptions. The main HARDI techniques currently existing are:

Higher order tensors [Des06]. The two–dimensional Cartesian tensor $\mathscr{D}$ may be easily generalized to an N–dimensional matrix, so this was one of the first attempts to overcome the limitations of the tensor model. It has two main problems: first, it is intended to characterize the ADC $D(\mathbf{g})$ but not $P(\mathbf{R}|\Delta)$, so it drives to representations quite difficult to interpret. Second, it is able to characterize $D(\mathbf{g})$ only for the measured b–value, so the assumption that the ADC is nearly constant for all b–values underlies this model.

Multitensor models [Ber07a, Kre05, Pel06]. They assume that eq. (2.14) may be written as the mixture of N Gaussian distributions, and the fiber bundles in the extent of the voxel to be clustered in one of the N modes. Maximum Likelihood or Least Squares techniques may be used to infer the tensor parameters and partial volumes of each cluster. The main problems of the model are: first, the need to know the number N of clusters beforehand; second: a mixture of Gaussians is an oversimplified model for the complex micro–architectures of fiber bundles. Besides fiber crossings, situations like fiber bending or kissing may occur, for which a Gaussian approximation cannot be justified.

Spherical deconvolution [And05, Des09, Tou07, Tou08], which generalize multitensor models. In this case, it is assumed that a fiber bundle is present at each an every spatial direction $\mathbf{r}$ with a given partial volume fraction $\alpha(\mathbf{r})$. Given the response $E_0(\mathbf{q})$ produced by a fiber bundle at direction $\mathbf{r}$, the deconvolution kernel, the global response $E(\mathbf{q})$ will be the superposition of each individual response, pondered by its partial volume fraction. The main drawback of this technique is the need to estimate the deconvolution kernel, and besides the need to assume that this kernel is the same for all voxels, which is not very realistic. Additionally, deconvolution is an ill–posed problem, so a regularization scheme is needed.

Continuous mixtures of Gaussians [Jia07]. Once again, the aim is to generalize the multitensor model. Like in spherical deconvolution, it is assumed that the signal in each voxel is the result of the continuous superposition of individual, Gaussian distributed, elemental fiber bundles at each direction. The elemental fiber bundles are supposed to follow a prior Wishart distribution which has to be fitted to data, so the orientation information may be computed as the Laplace transform of $E(\mathbf{q})$. Obviously, the limitation of this approach is the need to assume a prior Wishart distribution.

Persistent angular structures [Jan03]. Instead of the marginalization of $P(\mathbf{R}|\Delta)$ in $\mathbf{R}$, it is assumed that the diffusion of water molecules is limited to a sphere of radius R_0 for all possible orientations $\mathbf{r}$. This allows to compute the orientation information from a unique b–value as an inverse problem, by means of maximum entropy constrained optimization. The obvious drawback is the highly unrealistic assumption of the model, together with a heavy computational load.

The Diffusion Orientation Transform (DOT) [Öza06]. Like implicitly in higher order tensors, it is assumed that the ADC is constant for all b–values. This is equivalent to eliminate the dependence of D with b in eq. (2.39); like in DTI, this allows to completely characterize $E(\mathbf{q})$ from one single b. This property, together with a Spherical Harmonics (SH) analysis [Fra02], is exploited to compute the probability profile $P(R_0\mathbf{r}|\Delta)$ for a given R_0. The main drawbacks are: first, the need to assume a very restrictive model for $E(\mathbf{q})$, and second, the orientation information computed, $P(R_0\mathbf{r}|\Delta)$, has not an strict probabilistic interpretation.

Q–Ball imaging [Tuc03, Tuc04]. This is probably the most elegant solution, since it does not require any assumption on the behavior of $E(\mathbf{q})$ for b–values different from the one measured. This is possible thanks to the use of the Funk–Radon transform (FRT) of the measured signal $E(\mathbf{q})$, which may be proved to be an accurate estimator of the ODF of $P(\mathbf{R}|\Delta)$. This technique has two main drawbacks: first, the ODF is not an orientation PDF in the strict sense, so no probabilistic information may be inferred. Second, the use of the Funk–Radon transform yields an undesired angular blurring in the ODF, whose main effect is the increase in the uncertainty of the localization of fiber bundles.

The previous paragraphs are intended to be only a brief review of the most important HARDI techniques in the recent literature. A detailed analysis of the approaches closest related to the work presented in this dissertation will be carried out in Chapters 6 and 7. However, this study stresses the interest of the formal characterization of the diffusion propagator beyond DTI, i.e. the importance of the problem of relating the **q**–space to the **R**–space introduced in Section 2.5.3. Indeed, the works mentioned above have widely reported the greater power of HARDI techniques to characterize neural architectures beyond the information DTI provides.

Many of the techniques surveyed in Setion 2.6.1 for conventional DTI have been generalized to more complex scenarios by means of HARDI. The scope of this dissertation lies within the estimation of diffusion propagators and the orientation information inferred from them, so the representation, interpretation, and higher–level applications derived from this information are off–topics. Even so, it is interesting to briefly summarize some of the possibilities HARDI analyses provide. For example, the MD in DTI is defined as the trace of the diffusion tensor (the summation of its eigenvalues), and represents the overall capability of water molecules to diffuse through the tissue being imaged. With HARDI data sets, this measurement can be extended to the generalized trace, defined as the average of the ADC for all possible orientations [Des06, Öza05]:

$$\text{gentr}(D(\mathbf{g})) = \frac{3}{4\pi}\int_{\mathscr{S}} D(\mathbf{g})d\mathbf{g}, \tag{2.40}$$

where $\mathscr{S}$ is the sphere of radius 1, used to denote the average in the orientations **g**. The factor 3 is introduced as an analogy with the MD, which is three times the average among the eigenvalues. The ADC represents the amount of diffusion for each direction, much like the eigenvalues of the diffusion tensor represent the amount of diffusion for the principal directions given by the eigenvectors. Hence, the interpretation of the generalized trace is exactly the same as that of the MD.

On the other hand, the FA in conventional DTI admits a simple interpretation in terms of the trace of the normalized, squared diffusion tensor, which suggests the definition of a generalized variability of the form [Öza05]:

$$\mathscr{V}(D(\mathbf{g})) = \frac{1}{3}\left(\text{gentr}\left(D_n^2(\mathbf{g})\right) - \frac{1}{3}\right), \text{ for } D_n(\mathbf{g}) = \frac{D(\mathbf{g})}{\text{gentr}\left(D(\mathbf{g})\right)}. \tag{2.41}$$

Since this value is usually unbounded, a normalization is required:

$$\text{GA}(D(\mathbf{g})) = 1 - \left(1 + (250\mathscr{V})^{e(\mathscr{V})}\right)^{-1}, \text{ for } e(\mathscr{V}) = 1 + (1 + 5000\mathscr{V})^{-1}, \tag{2.42}$$

which defines the Generalized Anisotropy (GA) which now ranges from 0 to 1. This parameter is a generalized measure of the variability of diffusion for each spatial direction. Given this interpretation, a much more direct extrapolation of the FA can be thought of: in fact, the FA is the variance of the eigenvalues of the diffusion tensor normalized by their mean squared value. Hence, like it is done for the generalized trace, the eigenvalues (amount of diffusion for each principal direction) can be substituted by the ADC (amount of diffusion for each measured direction) to define:

$$\text{GA}(D(\mathbf{g})) = \frac{\frac{1}{4\pi}\int_{\mathscr{S}} D^2(\mathbf{g})d\mathbf{g} - \left(\frac{1}{4\pi}\int_{\mathscr{S}} D(\mathbf{g})d\mathbf{g}\right)^2}{\frac{1}{4\pi}\int_{\mathscr{S}} D^2(\mathbf{g})d\mathbf{g}}, \tag{2.43}$$

and this value lies already in the range $[0,1]$ without the need of further regularization. The integrals in **g** of eqs. (2.40) and (2.43) are replaced by summations for the directions measured. In any case, these indices provide richer information than the FA: those voxels where fiber crossings are present will drive to planar or spherical tensors, thus producing similar eigenvalues and a reduced FA. But a small value of the FA does not correspond to a poorly structured tissue, with

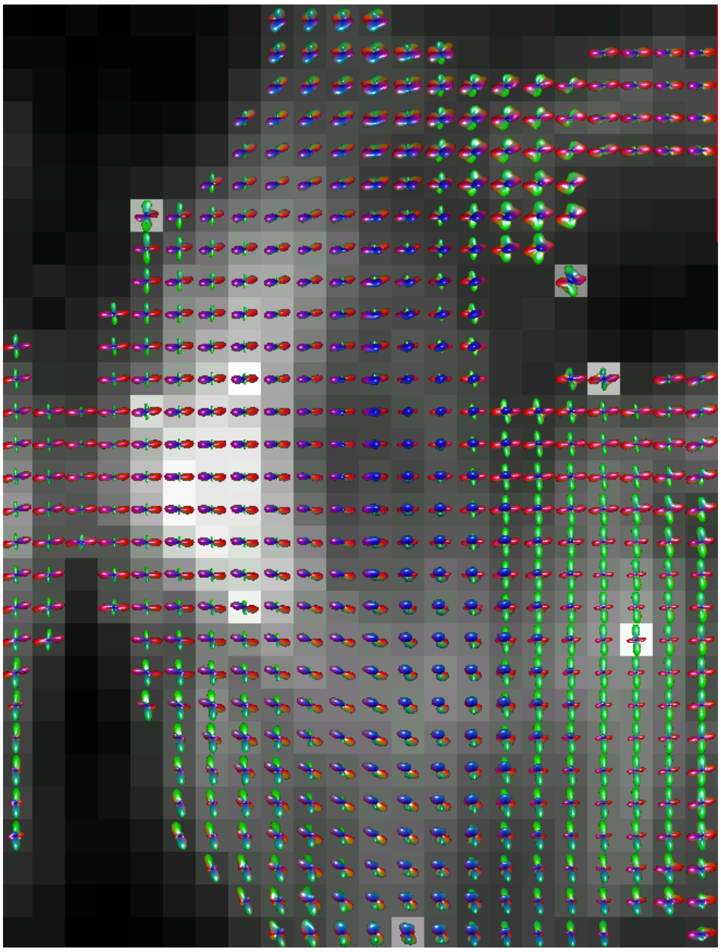


Figure 2.13: An example of glyph representation with HARDI, for an axial slice in the upper brain. Compared to the ellipsoids obtained with DTI, this kind of glyphs provides information on fiber crossings and complex neural architectures.

nearly isotropic diffusion but just the opposite. In these situations, the use of the GA can resolve this ambiguity.

Obviously, glyph representations may highly benefit from HARDI techniques, as suggested by Fig. 2.12. Indeed, this has been the preferred representation for diffusion propagators in the related literature [And05,Des09,Jan03,Öza06,Tou07, Tou08,Tuc03], and it is also the representation used for the results presented in this thesis. Fig 2.13 shows an illustration of the potential of these techniques, for an axial slice in the upper brain, where multiple fiber bundles of interest cross in all possible directions. Besides the glyph representation described in eq. (2.37), the background is colored according to the GA of the corresponding voxel, and glyphs are colored following color–coding conventions. This is a very representative example of how the combined use of the data provided by HARDI can offer highly valuable anatomical information in a very intuitive way.

Finally, as mentioned above, fiber tracking is the preferred application in the framework of Diffusion Imaging. The ability of HARDI techniques to resolve fiber crossings, bending, and other sources of ambiguity in the estimation of fiber populations is very useful for this task, and hence a number of recent efforts have been reported in this sense [Ber07b,Ber08,Des09,Zha09]. Still, it has been shown that tactography based on HARDI approaches can be used to elucidate neural architectures and connectivity which remain hindered in DTI studies [CA08].

3

Statistical characterization of noise in the $\mathbf{k}$–space and the $\mathbf{x}$–space

The present Chapter is focused on the study of the propagation of noise from the **k**–space to the **q**–space. For each gradient direction (each point in the **q**–space), an entire MRI acquisition has to be performed, so this is equivalent to study the propagation of noise from the **k**–space to the CMS. The noise in the **k**–space is mainly due to thermal disturbances in the MRI scanner, and its propagation to the **q**–space strongly depends on the algorithms used to reconstruct the DWI data set. The methods to reduce the amount of noise in this stage are beyond image processing, corresponding instead to hardware/scanning design. Consequently, this Chapter presents a statistical analysis of noise centered on existing techniques, but not new proposals to this respect. First, the statistics of noise in the **k**–space are studied: the common case of conventional MRI is generalized to multiple coils scanners. The study follows with the propagation of noise to the **x**–space by means of Fourier analysis. Finally, the Composite Magnitude Signal (CMS) has to be computed from complex signals computed from each receiving coil. In conventional MRI and most of pMRI algorithms, only one complex signal for all coils is available, so the Rician model holds. For multiple–coil MRI or the GRAPPA algorithm, one complex signal is available at each coil, so a more general model has to be used. The main contribution in this Chapter is in fact the study of such model. It is shown that under very weak assumptions GRAPPA signals can be modeled by means of non-central Chi statistics, which have been commonly used to described data sets acquired by phased arrays. Although the variance of noise may vary across different locations of the image domain, it is shown that a stationary model can be assumed in most of cases. Besides, it is argued that other pMRI techniques, such as SENSE, has to show this same feature, so image processing algorithms based on neighborhood operators apply as well for pMRI, since the statistics of noise are the same for all voxels inside the vicinity.

3.1 Introduction

As stated in the previous Chapter, the starting point for the image processing procedures in Diffusion Imaging are Diffusion Weighted Images (DWI). The methods presented in this dissertation may be seen as a processing pipeline, whose input are the DWI comprising the Composite Magnitude Signal (CMS) for each gradient image and baseline. Hence, no further processing is to be carried out in the raw images corresponding to the **k**–space or the complex **x**–space. However, MRI scanners and reconstruction protocols perform a number of operations in these domains that need to be analyzed to properly characterize the CMS, which is the aim of the present Chapter. In the simplest case of conventional acquisitions with one single receiving coil, the complex **x**–space is reconstructed from the **k**–space by means of Fourier transforming, thus carrying out purely linear operations. The noise can be assumed to be Gaussian in the complex **x**–space, but once the modulus of this signal is taken to compute the CMS, the statistics have to be modeled as Rician [Gud95], which highly complicates the analysis of DWI data sets. Nonetheless, the Rician model has been the keystone for most of the processing techniques in MRI, including noise filtering [AF08c, Bas06, Cou08, MF09, TV08], noise estimation [AF08a], and diffusion tensor estimation [Sal05], which stresses the paramount importance of the statistical characterization of noise.

With multiple coils systems or pMRI protocols, the modeling becomes more problematic. Coils systems with multiple channels were developed to enhance the SNR of the CMS while maintaining a large Field of View (FOV) [Con97][1]. If each coil is used to acquire the whole **k**–space independently, the SNR may be improved by simply averaging the multiple acquisitions. Yet, the pMRI techniques described in Chapter 2 extend the applicability of these systems by increasing the acquisition rate via subsampled takes of the **k**-space. This is especially important in Diffusion Imaging, where large amounts of data have to be acquired to characterize the **q**–space. In this case, each coil acquires a part of the **k**–space in parallel, and a reconstruction algorithm has to be used to gather these data and form a complete **k**–space. Dominant among these techniques are SENSE [Pru99] and GRAPPA [Gri02], and so the discussion will be centered on them.

Opposite to their original conception, multiple–coil systems for pMRI are not used to improve the SNR. Still, it is known that the SNR is drastically reduced by the use of these protocols due to two factors: first, to achieve a speedup in the acquisition time, the temporal averaging performed by the DFT is reduced; as a consequence, the noise power is increased proportionally to the square root of the speedup. Second, the reconstruction scheme, often based on a linear combination of measured **k**–space data, introduces an amplification of the noise level. This amplification has been characterized in SENSE reconstructions through the g–factor; although the interpretation of this parameter is not obvious for other pMRI algorithms, all of them suffer from this same artifact [Lar07]. The g–factor is used to indicate an intrinsic limit to the speedup of a given reconstruction scheme and receiver coil configuration. An additional difficulty related to the g–factor is that it is spatially variant, since it depends on the location of the coils employed. So, the noise power varies across the image.

From Section 2.4.1, a fundamental difference between the pMRI algorithms reviewed can be recognized: SENSE performs the reconstruction in the image domain, and so a unique complex **x**–space is available. Given this property, the CMS is computed as the modulus of a Gaussian distributed complex variable, and therefore SENSE data sets may be related to conventional MRI under a statistical point of view. A Rician behavior can indeed be assumed, and all the existing techniques for processing of conventional MRI can be used whenever the obstacle of the spatially variant noise power may be somehow overcome. On the contrary, with GRAPPA, a complete **k**–space, and consequently a complex image domain, is reconstructed for each receiving coil. This feature assimilates GRAPPA to multiple coils systems without subsampling. Concerning the characterization of DWI, this scenario drives to a non–central Chi model [Con97] if no subsampling is used and the CMS is computed by means of Sum of Squares (SoS). For GRAPPA, not only the noise variance is different at each image location, but also for each receiving coil at the same voxel. Although there exist some preliminary studies [Thu07], and even the non–central Chi model has been previously used without further justification [Die08], it is shown in this Chapter that the formal characterization of GRAPPA signals by means of the non–central Chi model is not so trivial in general, due to the different variance at each coil.

---

[1] Increasing the FOV requires sampling the **k**–space more densely to avoid aliasing artifacts. As a consequence, the acquisition has to be accelerated by decreasing the number of repetitions (NEX), the relaxation time, or both, thus worsening the SNR.

Consequently, the most challenging problem in the statistical characterization of the CMS is in GRAPPA reconstructed signals, since theoretical foundations for all the other common scenarios exist. GRAPPA reconstruction is entirely performed in the **k**–space, as opposed to SENSE, for which the reconstruction is performed in the image domain and does not need to interpolate the **k**–space. Nonetheless, SENSE can be mapped to a **k**-space reconstruction yielding the mSENSE method [Wan01], so in fact the results presented for GRAPPA may typically be generalized to other pMRI reconstruction methods as well. Thus, the study in this Chapter will be mainly focused on this algorithm, although a survey of known statistical properties of MRI data sets is provided as well for the sake of completeness. In particular, the following problems are addressed:

1. Statistical characterization of noise in the **k**–space. For conventional acquisitions, points in the **k**–space are directly provided by the receiving coils as RF signals with constant noise power. For parallel acquisitions, the characterization depends on the algorithm used. In particular, GRAPPA performs an interpolation of the **k**–space, giving rise to a variable noise power depending on wether the line has been directly acquired or it has been interpolated.
2. Statistical characterization of noise in the complex **x**–space. For conventional acquisitions, this signal is related to that in the previous stage by the DFT. For parallel acquisitions, the relation depends on the particular algorithm used. For GRAPPA, since the interpolation is carried out in the **k**–space, this same Fourier relation holds.
3. Statistical characterization of the CMS. The cases to contemplate are wether one single complex **x**–space (conventional MRI and SENSE acquisitions) or multiple spaces (multiple coils without subsampling and GRAPPA) are available. Besides, for GRAPPA it is necessary to analyze the conditions to fulfill for the non–central Chi model to hold.

3.2 Noise statistics in the k–space

3.2.1 Non–parallel MRI and pMRI without k–space sub–sampling

The physical principle driving the formation of the **k**–space from the phase/frequency encoding sequences was given in eq. (2.4), which is reproduced here for convenience:

$$s_l(\mathbf{k}) = \int_V W_l(\mathbf{x})\rho(\mathbf{x}) \exp\left(j2\pi\mathbf{k}^T\mathbf{x}\right) d\mathbf{x}, \tag{3.1}$$

where $\rho(\mathbf{x})$ is the (weighted) excited spin density function throughout the volume V, $\mathbf{x}$ is the spatial position within the FOV, and $W_l(\mathbf{x})$ is the spatial sensitivity of the receiving coil l at $\mathbf{x}$. Compared to eq. (2.4), an index l has been included to account for the possibility of multiple receiving coils, $0 \le l < L$. In conventional MRI, $L = 1$. To acquire each complete line in the **k**–space, the sequences are repeated with a different phase encoding each time. For each line, the RF signal is uniformly sampled at the desired rate. As a consequence, points in the **k**–space provided by the MRI scanner are independent samples measured from an RF signal, for which the noise can be modeled as a complex stationary Additive White Gaussian Noise (AWGN) process, with zero mean and variance σ_K^2. From a merely statistical point of view, the acquired signal $s_l(\mathbf{k})$ in eq. (3.1) may thus be modeled as:

$$s_l(\mathbf{k}) = a_l(\mathbf{k}) + N_l(\sigma_K^2), \tag{3.2}$$

with $a_l(\mathbf{k})$ the noise–free signal, $N_l(\sigma_K^2) = N_{l_c}(\mathbf{k};\sigma_K^2) + jN_{l_s}(\mathbf{k};\sigma_K^2)$, and N_{l_c} and N_{l_s} independent AWGN processes with variance σ_K^2.

3.2.2 pMRI with $\mathbf{k}$–space sub–sampling

As stated before, GRAPPA reconstruction is based on the interpolation of the whole $\mathbf{k}$–space at each coil out from a subsampled $\mathbf{k}$–space [Bla04,Gri02,Hog05]. On one hand, $N_l(\sigma_K^2)$ is no longer white noise, since the linear interpolations to estimate missing lines introduce strong correlations between measured and interpolated lines. On the other, this process affects as well the stationarity of the noisy processes: missing samples are computed as a linear combination of measured ones with noise power σ_K^2, so the resulting noise power will be σ_K^2 pondered by the mean squared value of the coefficients ω_{lj} used for the interpolation.

The details on the derivation of GRAPPA reconstruction were reviewed in Section 2.4.1). While the sampled lines (those directly provided by the MRI scanner), $s_l^{\mathscr{S}}(\mathbf{k})$, remain the same, the reconstructed (interpolated) lines at each coil, $s_l^{\mathscr{R}}(\mathbf{k})$, are estimated through a linear combination of existing samples, both from the same and from different coils. Weighted data in a neighborhood $\eta(\mathbf{k})$ around the estimated pixel from several coils are used for such estimation:

$$s_l^{\mathscr{R}}(\mathbf{k}) = \sum_{m=1}^{L} \sum_{\mathbf{c}\in\eta(\mathbf{k})} s_m^{\mathscr{S}}(\mathbf{k}-\mathbf{c})\omega_m(l,\mathbf{c},\mathbf{k}), \tag{3.3}$$

with $\omega_m(l,\mathbf{c},\mathbf{k}) = \alpha_m(l,\mathbf{c},\mathbf{k}) = +j\beta_m(l,\mathbf{c},\mathbf{k})$ the complex reconstruction coefficients for coil l. In self-referenced reconstructions, these coefficients are determined from the low-frequency coordinates of the $\mathbf{k}$-space, the so-called Auto Calibrating Signal (ACS) lines, which are sampled at the Nyquist rate (i.e. unaccelerated). Although the coefficients can vary across $\mathbf{k}$, this is not the most usual case, so the coefficients will be considered to be the same for all $\mathbf{k}$. Yet, they are in general different for each receiving coil, i.e. $\omega_m(l,\mathbf{c},\mathbf{k}) = \omega_m(l,\mathbf{c})$. The reconstructed lines are estimated linearly, and consequently the resulting data at each coil will match the following model:

$$s_l(\mathbf{k}) = \widehat{a}_l(\mathbf{k}) + N_l(\mathbf{k};\sigma_K^2(l,\mathbf{k})), \tag{3.4}$$

with $\widehat{a}_l(\mathbf{k})$ the (estimated) signal in the absence of noise, which has to be interpolated if $\mathbf{k}$ corresponds to an unsampled line. $N_l(\mathbf{k};\sigma_K^2(l,\mathbf{k}))$ are Gaussian noises with variance $\sigma_K^2(l,\mathbf{k})$ which now does depend on $\mathbf{k}$: the noise in the interpolated lines will be the weighted sum of different Gaussian noise realizations, so the noise in the final signal (in the $\mathbf{k}$–space) will also be Gaussian with zero mean, but neither stationary nor uncorrelated. For interpolated samples, the noise power will be, for the real part (the noise power of the imaginary part can be computed in the same way, and with the same result):

$$\begin{aligned}
\sigma_K^2(l,\mathbf{k}) &= E\left\{\Re\left\{s_l^{\mathscr{R}}(\mathbf{k}) - E\{s_l^{\mathscr{R}}(\mathbf{k})\}\right\}^2\right\} = E\left\{\Re\left\{N_l^{\mathscr{R}}(\mathbf{k};\sigma_K^2(l,\mathbf{k}))\right\}^2\right\} \\
&= E\left\{\sum_{m=1}^{L}\sum_{\mathbf{c}\in\eta(\mathbf{k})}\left(N_{m_c}^{\mathscr{S}}(\mathbf{k}-\mathbf{c})\alpha_m(l,\mathbf{c},\mathbf{k}) - N_{m_s}^{\mathscr{S}}(\mathbf{k}-\mathbf{c})\beta_m(l,\mathbf{c},\mathbf{k})\right)^2\right\} \\
&= \sum_{m=1}^{L}\sum_{\mathbf{c}\in\eta(\mathbf{k})} E\left\{\left(N_{m_c}^{\mathscr{S}}(\mathbf{k}-\mathbf{c})\right)^2\right\}\alpha_m^2(l,\mathbf{c},\mathbf{k}) + E\left\{\left(N_{m_s}^{\mathscr{S}}(\mathbf{k}-\mathbf{c})\right)^2\right\}\beta_m^2(l,\mathbf{c},\mathbf{k}) \\
&= \sigma_K^2\sum_{m=1}^{L}\sum_{\mathbf{c}\in\eta(\mathbf{k})}|\omega_m(l,\mathbf{c},\mathbf{k})|^2,
\end{aligned} \tag{3.5}$$

where ω^* denotes the complex conjugate of ω, and the independence between sampled values (and real and imaginary parts) has been exploited to compute the expectation of the modulus of the noisy process. Summarizing this Section, sampled lines have a constant noise power σ_K^2 determined by thermal disturbances of the scanner. Interpolated lines, on the contrary, have a noise power dependent on their **k**–space location and receiving coil l. Besides, the noise **is not** uncorrelated, neither between positions in **k**, nor from coil to coil.

3.3 Noise statistics in the complex x–space

3.3.1 Non–parallel MRI and pMRI without k–space sub–sampling

The complex **x**–space is obtained from the **k**–space by computing the 2D iDFT of $s_l(\mathbf{k})$ for each slice. Since the iDFT is a linear operator, the noise in the complex signal in the **x**–space for each coil (if more than one is used) will also be Gaussian, and it can be expressed as:

$$S_l(\mathbf{x}) = A_l(\mathbf{x}) + n_l(\mathbf{x};\sigma_n^2), \tag{3.6}$$

where $n_l(\mathbf{x};\sigma_n^2) = n_{l_c}(\mathbf{x};\sigma_n^2) + jn_{l_s}(\mathbf{x};\sigma_n^2)$ is a complex Gaussian process with zero mean and variance $\sigma_n^2(\mathbf{x})$. From eq. (3.2), the variance $\sigma_n^2(\mathbf{x})$ of $n_l(\mathbf{x};\sigma_n^2)$ may be easily computed following a similar reasoning to that in eq. (3.5). Consider that P samples have been measured for the first component of **k** (the frequency encoding direction), and Q samples for the second component (phase encoding). Consider as well the following matrix:

$$\mathscr{Q} = \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}, \tag{3.7}$$

whose determinant $\det(\mathscr{Q}) = PQ$ is the number of samples in the domain Ω of the image, $|\Omega| = \det(\mathscr{Q}) = PQ$. The noise power in the iDFT (for the real part) may be expressed in the form:

$$\begin{aligned} \sigma_n^2(\mathbf{x}) &= E\left\{\Re\left\{n_l(\mathbf{x};\sigma_n^2)\right\}^2\right\} = E\left\{\Re\left\{\frac{1}{|\Omega|}\sum_{\mathbf{k}\in\Omega} N_l(\mathbf{k};\sigma_K^2)\exp\left(j2\pi\mathbf{k}^T\mathscr{Q}^{-1}\mathbf{x}\right)\right\}^2\right\} \\ &= \frac{1}{|\Omega|^2}\sum_{\mathbf{k}\in\Omega}\sigma_K^2\left|\exp\left(j2\pi\mathbf{k}^T\mathscr{Q}^{-1}\mathbf{x}\right)\right|^2 = \frac{1}{|\Omega|^2}\sum_{\mathbf{k}\in\Omega}\sigma_K^2 = \frac{\sigma_K^2}{|\Omega|}. \end{aligned} \tag{3.8}$$

In fact, the noise power σ_n^2 is the same for all **x** within the FOV, so that n_l is a stationary process. Note that all samples in the **k**–space are combined in a weighted average to compute each sample in the **x**–space. Even so, it is easy to prove that the noise in the complex **x**–space is uncorrelated:

$$\begin{aligned} & E\{n_l(\mathbf{x};\sigma_n^2)n_l^*(\mathbf{x}';\sigma_n^2)\} \\ = {} & E\left\{\frac{1}{|\Omega|^2}\sum_{\mathbf{k}\in\Omega} N_l(\mathbf{k};\sigma_K^2)\exp\left(j2\pi\mathbf{k}^T\mathscr{Q}^{-1}\mathbf{x}\right)\sum_{\mathbf{l}\in\Omega} N_l^*(\mathbf{l};\sigma_K^2)\exp\left(-j2\pi\mathbf{l}^T\mathscr{Q}^{-1}\mathbf{x}'\right)\right\} \\ = {} & \frac{1}{|\Omega|^2}\sum_{\mathbf{k}\in\Omega}\sum_{\mathbf{l}\in\Omega} E\left\{N_l(\mathbf{k};\sigma_K^2)N_l^*(\mathbf{l};\sigma_K^2)\right\}\exp\left(j2\pi\left(\mathbf{k}^T\mathscr{Q}^{-1}\mathbf{x} - \mathbf{l}^T\mathscr{Q}^{-1}\mathbf{x}'\right)\right) \\ = {} & \frac{1}{|\Omega|^2}\sum_{\mathbf{k}\in\Omega} 2\sigma_K^2\exp\left(j2\pi\mathbf{k}^T\mathscr{Q}^{-1}(\mathbf{x}-\mathbf{x}')\right) \\ = {} & \frac{2\sigma_K^2}{|\Omega|^2}\sum_{k_f=0}^{P-1}\exp\left(j2\pi\frac{k_f(x_f-x_f')}{P}\right)\sum_{k_p=0}^{Q-1}\exp\left(j2\pi\frac{k_p(x_p-x_p')}{Q}\right), \end{aligned} \tag{3.9}$$

where the notation $\mathbf{k} = [k_f, k_p]^T$, and $\mathbf{x} = [x_f, x_p]^T$ has been used, and it has been exploited that N_l are uncorrelated for different wave numbers **k**. In the previous equation, if $\mathbf{x} = \mathbf{x}'$, the arguments of the exponentials are 0, so all terms in the summations are simply 1 and the expression reduces to twice the variance of noise in the **x**–space (since both the real and imaginary parts are considered in the previous development). For $\mathbf{x} \neq \mathbf{x}'$, it is easy to check that both summations are null; for the first one:

$$\sum_{k_f=0}^{P-1} \exp\left(j2\pi \frac{k_f(x_f - x'_f)}{P}\right) = \frac{\exp\left(j2\pi \frac{P(x_f-x'_f)}{P}\right) - 1}{\exp\left(j2\pi \frac{(x_f-x'_f)}{P}\right) - 1} = \frac{1-1}{\exp\left(j2\pi \frac{(x_f-x'_f)}{P}\right) - 1} = 0, \quad (3.10)$$

since eq. (3.10) is the sum of a geometric progression, and $x_f - x'_f$ is necessarily an integer. Summarizing this Section, the noise in the complex **x**–space, for conventional acquisitions, can be modeled as a stationary AWGN process with zero mean and variance $\sigma_n^2 = \sigma_K^2/|\Omega|$. For multiple coils systems without **k**–space subsampling, this same model holds for each coil.

3.3.2 pMRI with k–space sub–sampling

An analogous development to that in eq. (3.8) is not so trivial for GRAPPA reconstructions. First, the noise variance is different at each location in the **k**–space, and second, the noisy samples are not uncorrelated after interpolation, so the expectation of the modulus of the summation is not equal to the summation of the moduli of each term. The expression of the noisy process, of course, is the same as in the previous case, involving the iDFT summation. Consequently, the noise in the complex image domain is still a Gaussian process with zero mean. It is convenient to separate those terms depending on sampled lines from those depending on interpolated ones:

$$\begin{aligned} n_l(\mathbf{x};\sigma_n^2) &= \frac{1}{|\Omega|}\sum_{\mathbf{k}\in\Omega} N_l(\mathbf{k})\exp\left(j2\pi\mathbf{k}^T\mathscr{Q}^{-1}\mathbf{x}\right) \\ &= \frac{1}{|\Omega|}\left(\sum_{\mathbf{k}\in\mathscr{S}} N_l^{\mathscr{S}}(\mathbf{k})\exp\left(j2\pi\mathbf{k}^T\mathscr{Q}^{-1}\mathbf{x}\right) + \sum_{\mathbf{k}'\in\mathscr{R}} N_l^{\mathscr{R}}(\mathbf{k}')\exp\left(j2\pi\mathbf{k}'^T\mathscr{Q}^{-1}\mathbf{x}\right)\right), \end{aligned} \quad (3.11)$$

where the dependency with $\sigma_K^2(l,\mathbf{k})$ in N_l has been dropped for the sake of clarity. All samples corresponding to the subset $\mathscr{S} \subset \Omega$ are independent, since they correspond to values actually acquired by the MRI scanner. On the contrary, samples in $\mathscr{R} \subset \Omega$ are linear combinations of samples in $\mathscr{S}$, so independence cannot be assumed. These lines are obtained as described in eq (3.3), which can be used to derive the form of the interpolated noisy processes:

$$\begin{aligned} N_l^{\mathscr{R}}(\mathbf{k}) &= \sum_{m=1}^{L}\sum_{\mathbf{c}\in\eta(\mathbf{k})} N_m^{\mathscr{S}}(\mathbf{k}-\mathbf{c})\omega_m(l,\mathbf{c}) \\ &= \sum_{\mathbf{c}\in\eta(\mathbf{k})} N_l^{\mathscr{S}}(\mathbf{k}-\mathbf{c})\omega_l(l,\mathbf{c}) + \sum_{m\neq l}\sum_{\mathbf{c}\in\eta(\mathbf{k})} N_m^{\mathscr{S}}(\mathbf{k}-\mathbf{c})\omega_m(l,\mathbf{c}). \end{aligned} \quad (3.12)$$

Sampled values have been separated in two different sets for convenience: one corresponds to samples in the same coil as the sample to interpolate, and the other to samples in the remaining coils. Besides, it is convenient to distinguish between two different subsets in $\mathscr{S}$: the ACS lines, belonging to subset $\mathscr{A}$, are not used to interpolate any line inside $\mathscr{R}$. A new subset $\mathscr{S}' \subset \mathscr{S} \subset \Omega$ is introduced such that $\mathscr{S} = \mathscr{A} \cup \mathscr{S}'$, denoting those lines in $\mathscr{S}$ which are actually used to interpolate lines in $\mathscr{R}$. With this notation, eq. (3.12) can now be casted into eq. (3.11) to yield:

$$\begin{aligned} n_l(\mathbf{x};\sigma_n^2) \;=\; & \frac{1}{|\Omega|}\sum_{\mathbf{k}\in\mathscr{S}'}\left(N_l^{\mathscr{S}'}(\mathbf{k})\left(\exp\left(j2\pi\mathbf{k}^T\mathscr{Q}^{-1}\mathbf{x}\right)+\sum_{\mathbf{c}\in\eta(\mathbf{k})}\omega_l(l,\mathbf{c})\exp\left(j2\pi(\mathbf{k}+\mathbf{c})^T\mathscr{Q}^{-1}\mathbf{x}\right)\right)\right. \\ & \left.+\sum_{m\neq l}N_m^{\mathscr{S}'}(\mathbf{k})\sum_{\mathbf{c}\in\eta(\mathbf{k})}\omega_m(l,\mathbf{c})\exp\left(j2\pi(\mathbf{k}+\mathbf{c})^T\mathscr{Q}^{-1}\mathbf{x}\right)\right) \\ & +\frac{1}{|\Omega|}\sum_{\mathbf{k}\in\mathscr{A}}N_l^{\mathscr{A}}(\mathbf{k})\exp\left(j2\pi\mathbf{k}^T\mathscr{Q}^{-1}\mathbf{x}\right). \end{aligned} \tag{3.13}$$

In the previous equation, only values actually sampled have been used, so independence between them can now be assumed. Once again, the linearity of the operations in the previous equation assures that the noise in the complex **x**–space is Gaussian. Since it has zero mean, its variance equals its mean squared value. The variance of the real part of eq. (3.13), $E\{n_{l_c}^2\}$, is computed next (for the imaginary part, exactly the same derivation can be followed). To simplify the expressions, the notation $\widetilde{\mathbf{x}}=2\pi\mathscr{Q}^{-1}\mathbf{x}$ is used in what follows. The variance is computed as:

$$\begin{aligned} E\{n_{l_c}^2\} \;=\; & \frac{1}{|\Omega|^2}\sum_{\mathbf{k}\in\mathscr{S}'}\left[\left(\cos(\mathbf{k}^T\widetilde{\mathbf{x}})+\sum_{\mathbf{c}\in\eta(\mathbf{k})}\alpha_l(l,\mathbf{c})\cos((\mathbf{k}+\mathbf{c})^T\widetilde{\mathbf{x}})\right.\right. \\ & \left.-\sum_{\mathbf{c}\in\eta(\mathbf{k})}\beta_l(l,\mathbf{c})\sin((\mathbf{k}+\mathbf{c})^T\widetilde{\mathbf{x}})\right)^2 E\{N_{l_c}^2(\mathbf{k})\}+\left(\sin(\mathbf{k}^T\widetilde{\mathbf{x}})\right. \\ & \left.+\sum_{\mathbf{c}\in\eta(\mathbf{k})}\alpha_l(l,\mathbf{c})\sin((\mathbf{k}+\mathbf{c})^T\widetilde{\mathbf{x}})+\sum_{\mathbf{c}\in\eta(\mathbf{k})}\beta_l(l,\mathbf{c})\cos((\mathbf{k}+\mathbf{c})^T\widetilde{\mathbf{x}})\right)^2 E\{N_{l_s}^2(\mathbf{k})\} \\ & +\sum_{m\neq l}\left(\left(\sum_{\mathbf{c}\in\eta(\mathbf{k})}\alpha_m(l,\mathbf{c})\cos((\mathbf{k}+\mathbf{c})^T\widetilde{\mathbf{x}})-\sum_{\mathbf{c}\in\eta(\mathbf{k})}\beta_m(l,\mathbf{c})\sin((\mathbf{k}+\mathbf{c})^T\widetilde{\mathbf{x}})\right)^2 E\{N_{l_c}^2(\mathbf{k})\}\right. \\ & \left.\left.+\left(\sum_{\mathbf{c}\in\eta(\mathbf{k})}\alpha_m(l,\mathbf{c})\sin((\mathbf{k}+\mathbf{c})^T\widetilde{\mathbf{x}})+\sum_{\mathbf{c}\in\eta(\mathbf{k})}\beta_m(l,\mathbf{c})\cos((\mathbf{k}+\mathbf{c})^T\widetilde{\mathbf{x}})\right)^2 E\{N_{l_s}^2(\mathbf{k})\}\right)\right] \\ & +\frac{1}{|\Omega|^2}\sum_{\mathbf{k}\in\mathscr{A}}\left(\cos^2(\mathbf{k}^T\widetilde{\mathbf{x}})\;E\{N_{l_c}^2(\mathbf{k})\}+\sin^2(\mathbf{k}^T\widetilde{\mathbf{x}})\;E\{N_{l_s}^2(\mathbf{k})\}\right). \end{aligned} \tag{3.14}$$

The previous equation admits some simplifications taking into account that $E\{N_{l_c}^2\}=E\{N_{l_s}^2\}=\sigma_K^2$, since only sampled lines are considered. Even so, the resulting expression is far more complicated than the simple relation in eq. (3.8). The proof of the following result, although quite tedious, is straightforward from the previous development:

$$\sigma_{n_l}^2(\mathbf{x})=E\{n_{l_c}^2\}=E\{n_{l_s}^2\}=\frac{\sigma_K^2}{|\Omega|^2}\left(|\mathscr{S}'|\,(1+\Upsilon_l(\mathbf{x}))+|\mathscr{A}|+\sum_{\mathbf{k}\in\eta(\mathbf{k})}W_l(\mathbf{k})^2\right), \tag{3.15}$$

where, as always, $|\mathscr{S}'|$ and $|\mathscr{A}|$ denote the cardinals of subsets $\mathscr{S}'$ and $\mathscr{A}$. The functions Υ and W are defined for each coil as:

$$\begin{aligned}\Upsilon_l(\mathbf{x}) &= 2\sum_{\mathbf{c}\in\eta(\mathbf{k})}\left(\alpha_l(l,\mathbf{c})\cos(\mathbf{c}^T\widetilde{\mathbf{x}})-\beta_l(l,\mathbf{c})\sin(\mathbf{c}^T\widetilde{\mathbf{x}})\right)+\\ &\quad+\sum_{m=1}^{L}\left(\sum_{\mathbf{c}\neq\mathbf{q}}(\alpha_m(l,\mathbf{c})\alpha_m(l,\mathbf{q})-\beta_m(l,\mathbf{c})\beta_m(l,\mathbf{q}))\cos((\mathbf{c}-\mathbf{q})^T\widetilde{\mathbf{x}})\right.\\ &\quad\left.+2\sum_{\mathbf{c},\mathbf{q}}\alpha_m(l,\mathbf{c})\beta_m(l,\mathbf{q})\sin((\mathbf{c}-\mathbf{q})^T\widetilde{\mathbf{x}})\right); \qquad (3.16)\\ W_l(\mathbf{k})^2 &= \sum_{m=1}^{L}\sum_{\mathbf{c}\in\eta(\mathbf{k})}|\omega_m(l,\mathbf{c},\mathbf{k})|^2. \qquad (3.17)\end{aligned}$$

In the most common case, the reconstruction coefficients are equal for all wave numbers $\mathbf{k}$: $\omega_m(l,\mathbf{c},\mathbf{k})=\omega_m(l,\mathbf{c})$, and W_l does not depend on $\mathbf{k}$. Therefore, eq. (3.15) simplifies, without much loss of generality, to the following expression:

$$\sigma_{n_l}^2(\mathbf{x})=\frac{\sigma_K^2}{|\Omega|^2}\left(|\mathscr{S}'|\left(1+W_l^2+\Upsilon_l(\mathbf{x})\right)+|\mathscr{A}|\right). \qquad (3.18)$$

As opposed to the conventional case (without **k**–space subsampling), now the variance of noise does depend on the image location **x** thorough $\Upsilon(\mathbf{x})$. Noise in the complex **x**–space is no longer stationary, unless the spatial variability of $\Upsilon(\mathbf{x})$ is small enough. Even if this assumption does not hold, the result in eq. (3.18) gives a map of noise variances and allows to characterize the SNR of the MRI data set, much like the g–factor does for SENSE. Although it is widely known that the variance of noise is not uniform in pMRI recontructions (see [Thu07] for instance), no theoretical results have been given before to this respect for GRAPPA.

Finally, the interpolation in the **k**–space introduces a certain statistical correlation between the samples, so the noise is not AWGN. When the iDFT is used to recover the image domain, this correlation translates to the **x**–space, so it cannot be assumed either that the noise is white in the image domain, as opposed to the conventional case without subsampling.

3.4 Noise statistics in the Composite Magnitude Signal

3.4.1 Conventional MRI

As described in Section 2.5, it is enough to know the magnitude of the signal in the image domain to characterize the diffusion process. In the simplest scenario of conventional MRI with one coil, each location **x** is characterized by a complex Gaussian noise with variance $\sigma_n^2=\sigma_K^2/|\Omega|$. Consequently, the CMS is:

$$M(\mathbf{x})=\sqrt{S_c^2(\mathbf{x})+S_s^2(\mathbf{x})}=\sqrt{(A_c(\mathbf{x})+n_c(\mathbf{x};\sigma_n^2))^2+(A_s(\mathbf{x})+n_s(\mathbf{x};\sigma_n^2))^2}, \qquad (3.19)$$

where the subindex l has been deliberately dropped down since only one coil is being considered. Eq. (3.19) drives to the well–known Rician model traditionally used in MRI [Gud95]. The PDF of the measured value is expressed as [Dru93]:

$$f_M(t;A,\sigma_n)=\frac{t}{\sigma_n^2}\exp\left(-\frac{t^2+A^2}{2\sigma_n^2}\right)I_0\left(\frac{A\,t}{\sigma_n^2}\right)u(t), \qquad (3.20)$$

where A is the magnitude of the noise–free signal, $A=\sqrt{A_c^2+A_s^2}$, I_0 is the modified Bessel function of the first kind and index 0, and u is Heaviside's step function. Although the dependence with $\mathbf{x}$ has been dropped down for brevity, the noise–free signal A (and therefore M) depends on the location $\mathbf{x}$. For conventional MRI, eq. (3.20) holds at each image location with the same noise variance σ_n^2. Besides, since the complex Gaussian process is white, Rician distributed samples in the CMS are indeed independent. For SENSE acquisitions, the CMS is computed as well as the modulus of a complex Gaussian variable, so the Rician model still holds. It is easy to understand that this same consideration is valid for all pMRI algorithms for which one single complex image domain is recovered. Yet, with pMRI acquisitions, the noise variance is not the same for all image locations, and noisy samples are not independent in general.

3.4.2 Multiple coils MRI

For multiple coils MRI, several replicates of the complex image domain are available, and they have to be somehow combined to obtain the CMS. The most popular way to do so is the SoS method [Con97, Roe90]:

$$M_L(\mathbf{x})=\sqrt{\sum_{l=1}^{L}S_{l_c}^2(\mathbf{x})+S_{l_s}^2(\mathbf{x})}=\sqrt{\sum_{l=1}^{L}\left(A_{l_c}(\mathbf{x})+n_{l_c}(\mathbf{x};\sigma_n^2)\right)^2+\left(A_{l_s}(\mathbf{x})+n_{l_s}(\mathbf{x};\sigma_n^2)\right)^2}, \qquad (3.21)$$

where the notation M_L is used to make it explicit that the CMS is the combination of the signals from L receiving coils. In fact, SoS is not only an heuristic to combine the L signals; it has been proved in [Gil07] that it represents the optimal estimate, in the sense of LS, when the sensitivity of each receiving coil is unknown, and therefore has to be estimated from data.

In case no subsampling is performed in the **k**–space, the noise variance σ_n^2 is the same for all image locations, and also for all coils: $\sigma_n^2=\sigma_K^2/|\Omega|$. If the noises at each coil are considered independent (see Section 3.5 for a discussion on this topic), the CMS, $M_L(\mathbf{x})$, is described in terms of a non–central Chi distribution, with PDF [Abr72]:

$$f_{M_L}(t;A_L,\sigma_n,L)=\frac{A_L^{1-L}}{\sigma_n^2}t^L\exp\left(-\frac{t^2+A_L^2}{2\sigma_n^2}\right)I_{L-1}\left(\frac{A_L\,t}{\sigma_n^2}\right)u(t), \qquad (3.22)$$

where now I_{L-1} is the Bessel function of the first kind an order $L-1$. It is easy to verify that, for $L=1$, this PDF reduces to the Rician distribution in eq (3.20), since it is the same case as if one single receiving coil were used. Like in this former equation, M_L and A_L in eq. (3.22) depend on the spatial location $\mathbf{x}$. If no subsampling is performed, non–central Chi–distributed samples are over again independent from each other.

Nonetheless, the generalization to pMRI is not so easy in this case. From eq. (3.18), it remains clear that the noise variance in GRAPPA reconstructions depends on $\mathbf{x}$ through Υ_l. What is more important, such function has a different definition for each coil, and as a consequence the noise variance $\sigma_n^2(\mathbf{x},l)$ is different for each coil. The non–central Chi distribution can only be assumed if the variance (but not the mean value) for all coils is the same, so it is not valid for GRAPPA signals in principle, as opposed to what it has been stated in [Die08][2]. The remainder of this Section is centered on the conditions the complex signals at each coil must fulfill for the non–central Chi model to approximately hold. To that

---

[2] Yet, there is an additional artifact which complicates the modeling of the CMS: the non–central Chi model requires that the noisy processes at each coil are independent from each other, which might not be the case. This issue is discussed in Section 3.5.

end, the squared CMS has to be studied instead of the CMS itself, i.e. the random variable $X_L = M_L^2$ is defined:

$$X_L(\mathbf{x}) = \sum_{l=1}^{L} |S_l(\mathbf{x};\sigma_{n_l}^2)|^2 = \sum_{l=1}^{L} \left(S_{l_c}^2(\mathbf{x};\sigma_{n_l}^2) + S_{l_s}^2(\mathbf{x};\sigma_{n_l}^2) \right), \tag{3.23}$$

where the notation $\sigma_{n_l}^2$ is used instead of σ_n^2 to point out that the noise variance is different for each coil. In case all $\sigma_{n_l}^2$ are equal, i.e. $\sigma_{n_l}^2 = \sigma_n^2$, X_L follows a non-central Chi squared distribution, with PDF [Abr72]:

$$f_{X_L}(t;A_L,\sigma_n,L) = \frac{A_L^{1-L}}{2\sigma_n^2} t^{(L-1)/2} \exp\left(-\frac{t+A_L^2}{2\sigma_n^2}\right) I_{L-1}\left(\frac{A_L\sqrt{t}}{\sigma_n^2}\right) u(t), \tag{3.24}$$

which is closely related to the PDF of M_L in eq. (3.22). On the other hand, if the variance $\sigma_{n_l}^2$ at each coil is different, the distribution of X_L is no longer non-central Chi squared. For each coil, $X_l = |S_l(\sigma_{n_l}^2)|^2$ will follow a non-central Chi squared law with $L = 1$. Its characteristic function, defined as $G_{X_l}(w) = E\{\exp(jwX_l)\}$ [Pap91] (note the relation with the inverse Fourier transform of f_{X_l}), will be:

$$G_{X_l}(w) = \frac{1}{1-2jw\sigma_{n_l}^2} \exp\left(\frac{|A_l|^2 jw}{1-2jw\sigma_{n_l}^2}\right). \tag{3.25}$$

The characteristic function of X_L can be easily computed if independence between all X_l is assumed. The resulting characteristic function is the product of the individual ones for each X_l:

$$G_{X_L}(w) = \prod_{l=1}^{L} G_{X_l}(w) = \prod_{l=1}^{L} \frac{1}{1-2jw\sigma_{n_l}^2} \exp\left(\frac{jw|A_l|^2}{1-2jw\sigma_{n_l}^2}\right), \tag{3.26}$$

which for equal variances reduces to:

$$G_{X_L}(w) = \frac{1}{(1-2jw\sigma_n^2)^L} \exp\left(\frac{\sum_{l=1}^{L} jw|A_l|^2}{1-2jw\sigma_n^2}\right) = \frac{1}{(1-2jw\sigma_n^2)^L} \exp\left(\frac{jw|A_L|^2}{1-2jw\sigma_n^2}\right). \tag{3.27}$$

The previous expressions are very useful for the subsequent developments. The PDF of the resulting variable X_L for different $\sigma_{n_l}^2$ cannot be explicitly computed, but its characteristic function can in fact be given an analytical expression. Furthermore, the moments of this distribution can be related to the derivatives of its characteristic function evaluated at $w = 0$ [Pap91], which allows to compute the mean and variance of X_L:

$$E\{X_L\} = E\{M_L^2\} = \begin{cases} A_L^2 + 2\sigma_n^2 L, & \text{equal } \sigma_{n_l}^2 = \sigma_n^2; \\ A_L^2 + 2\overline{\sigma_n^2} L, & \text{different } \sigma_{n_l}^2; \end{cases} \tag{3.28}$$

$$\mathrm{Var}\{X_L\} = \mathrm{Var}\{M_L^2\} = \begin{cases} 4A_L^2\sigma_n^2 + 4L\sigma_n^4, & \text{equal } \sigma_{n_l}^2 = \sigma_n^2; \\ 4A_L^2\overline{\sigma_n^2} + 4L\overline{\sigma_n^2}^2, & \text{different } \sigma_{n_l}^2; \end{cases} \tag{3.29}$$

where:

$$\overline{\sigma_n^2} = \frac{1}{L}\sum_{l=1}^{L} \sigma_{n_l}^2 \tag{3.30}$$

is the mean noise power. In fact, with the previous definition, the mean of the distributions of X_L, either with the same σ_n^2 or with different $\sigma_{n_l}^2$, is exactly the same. The variance is not equal due to the second term in eq. (3.29). The equivalence (with respect to the first two moments) would be complete if the equality:

$$\overline{\sigma_n^4} = \overline{\sigma_n^2}^2 \Rightarrow \overline{\sigma_n^4} - \overline{\sigma_n^2}^2 = 0 \Rightarrow \mathrm{Var}\{\sigma_{n_l}^2\} = 0 \tag{3.31}$$

held, which is only possible if all $\sigma_{n_l}^2$ are equal. Nonetheless, the previous development suggests that the statistical characterization of X_L for the case of different $\sigma_{n_l}^2$, in case their variance is not too large, can be assimilated to a non–central Chi squared variable, with equivalent constant noise power $\overline{\sigma_n^2}$. Hence, the condition for the non–central Chi model to hold for GRAPPA reconstructed data sets is precisely the small variability of noise variances at each receiving coil. Besides, the previous development provides a valuable tool to validate such assertion. Given the uniqueness of the Fourier transform, two random variables have the same PDF (i.e. they are equally distributed) if and only if they share the same characteristic function. Comparing eqs. (3.26) and (3.27) through the definition of $\overline{\sigma_n^2}$ in eq. (3.30), the accuracy of the approximation may be tested.

As a summary of this Section, the non–central Chi model applies to GRAPPA reconstructed signals whenever the variances of noise for each coil are approximately the same. This result is the main contribution in this Chapter, and thus will be thoroughly validated in Section 3.6. Contrary to the case of multiple coils systems without **k**–space subsampling, the model for GRAPPA is not stationary in principle (see Section 3.3), and the noise is not uncorrelated either. Yet, it is empirically shown in Section 3.6 that in fact GRAPPA signals can be considered nearly stationary under very weak assumptions.

3.5 Applicability of the models proposed

The previous developments have been carried out assuming ideal conditions. In particular, for multiple coils systems, the noise at each coil has been considered independent from the noises at the remaining coils, which is a condition for the non–central Chi model to hold. However, it is known that correlations between coils in phased arrays do exist [Har92,Hay90,Red92]. Yet, the effect of such correlations is often left aside due to their minimal impact and practical considerations. In [Con97] it is stated that the outcome of noise dependencies is minimal over humans and phantom data, although a bias in SNR estimation may appear.

For GRAPPA reconstructed signals, there is still another source of correlation. Even if the signals acquired by each receiving coil are independent, missing samples in the **k**–space are interpolated at each coil taking into account samples from all the remaining coils, which obviously introduces an additional interdependency between the noises at each reconstructed **x**–space. This artifact could alter the statistical characterization of the CMS following a non–central Chi distribution. However, it is considered in what follows that the overall correlation between the **x**–spaces may be kept low enough to guarantee the accuracy of the model described. As an additional remark, the study of correlation–related artifacts is still a current matter of study [Bro04], and it will be considered beyond the scope of this dissertation.

As a final comment, the assumption of independence between acquired samples in single coil systems may be unfulfilled in some situations. In Echo Planar Imaging (EPI) data sets, for example, not only the RF pulses, but also their echoes, are received and interpreted by the MRI scanner, so the acquisitions of different samples at the **k**–space are not completely independent [Pos95]. This artifact may introduce noticeable spatial correlations in the complex image domain, so that the noise is no longer AWGN. However, neither the Rician character of first order statistics, nor the stationary behavior of noise is altered in this case. Nonetheless, neighboring spatial samples are not independent with these protocols, which can affect some image processing algorithms based on local neighborhoods, like those

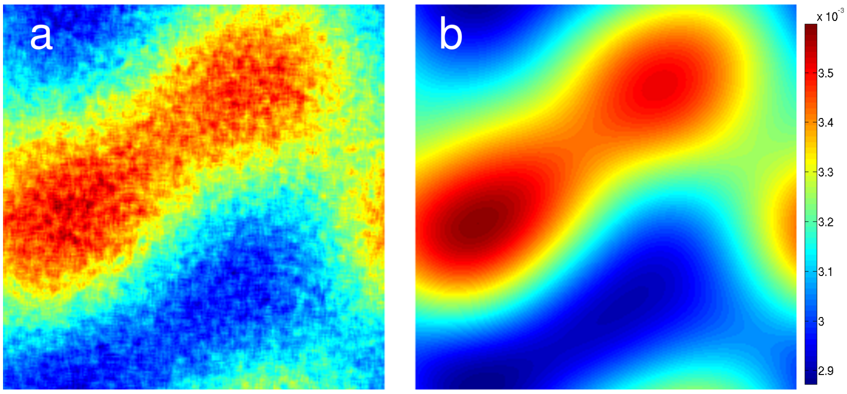


Figure 3.1: Map of variances in the $\mathbf{x}$–space for a synthetic image reconstructed with GRAPPA, for the first of 8 receiving coils. In (a), the variance has been directly estimated from local variances averaged for 1000 Montecarlo trials. In (b), the theoretical expression given in eq. (3.18) has been evaluated for the first trial.

described in Chapter 5 for noise removal.

3.6 Validation of the statistical characterization

This Section is devoted to validate the accuracy of the non–central Chi model in GRAPPA reconstructions. First, the fluctuation of the noise variance along each coil in the $\mathbf{x}$–space is analyzed through the use of synthetic data. Noise is simulated in the $\mathbf{k}$–space in an 8-coil system, by generating eight 256×256 complex Gaussian noisy images $N_l(\mathbf{k};\sigma_K^2)$, with zero mean and equal standard deviation $\sigma_K = 20$. The data is decimated using a $2\times$ line–subsampling of the $\mathbf{k}$-space, while 13 $1\times$ ACS lines are used near the origin of the $\mathbf{k}$–space. Missing lines are then interpolated using GRAPPA, and the iDFT is computed to reconstruct the $\mathbf{x}$–space corresponding to each coil. This experiment is repeated for 1000 realizations, keeping the reconstruction coefficients, $\omega_m(l,\mathbf{c})$, of the first realization for the sake of statistical validity. The theoretical variance is calculated using eq. (3.18), and the experimental variance is estimated at each spatial location as the sample variance in a local neighborhood of 5×5 pixels. This value is further averaged for the 1000 Montecarlo trials. Fig. 3.1 shows the results for the estimated variance and for the theoretical one; only the results for the first coil are shown. Obviously, the theoretical expression accurately describes the variance derived from GRAPPA reconstruction. In addition, the fluctuation of the variance across the image is not too high (it ranges from 2.9×10^{-3} to 3.6×10^{-3}). Yet, it is quite slow: considering nearby pixels in a spatial sense, the variance of noise is approximately constant. This consideration justifies the computation of the map in Fig. 3.1 by means of neighborhoods operations. Besides, it partially assures the consistency of image processing techniques, such as denoising filters, based on neighborhood operations. Nevertheless, the fluctuation of noise variance across the image is still too large to consider a nearly stationary behavior of noise.

The previous study suffers from two limitations: first, no signal is added to the noise in the $\mathbf{k}$–space, which could drive to unrealistic values of the weights $\omega_m(l,\mathbf{c})$ and thus of $\Upsilon(\mathbf{x})$[3]. Second, the noisy signals for each coil are generated without taking into account the sensitivity of the receiving coils. To overcome them, a number of experiments with real data sets BWH4, PULSAR1, and PULSAR2 (see

---

[3] Since the signal considered is pure noise, no correlations between lines should be present, so the LS fitting of interpolation weights is not significant. However, this first experiment aims only to check the accuracy of the analytical expressions obtained, and hence this is not an issue.

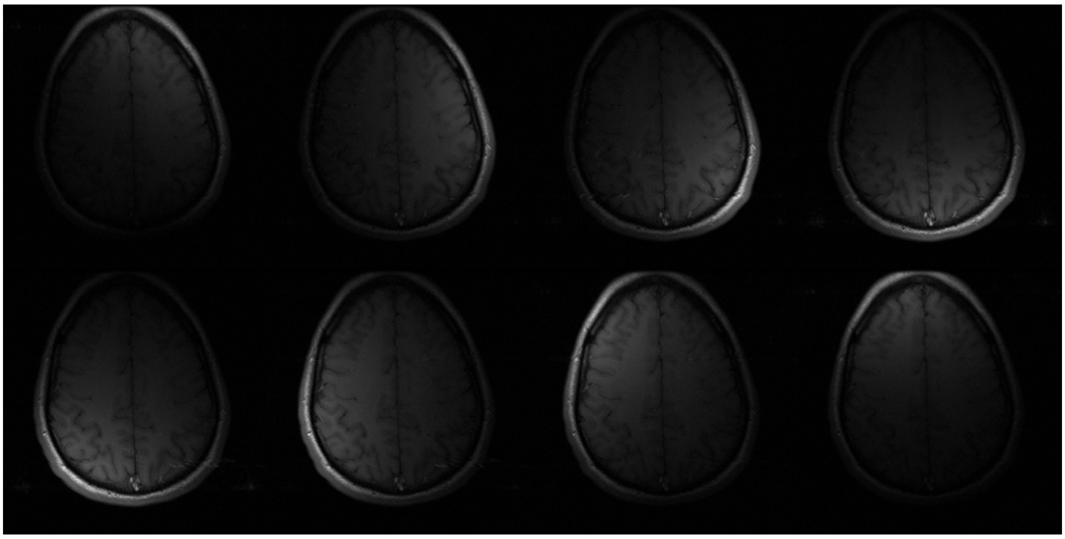


Figure 3.2: Data set BWH4: an axial slice in the upper brain, for each receiving coil in the 8 coils system.

Section 1.4) are presented next. Figs. 3.2 to 3.4 show respective slices of these volumes for all receiving coils available. For all these data sets, the different spatial sensitivity of each receiving coil is perfectly noticeable. The brightness of those areas of the image corresponding to the strongest sensitivities is much higher than the one of those with the lowest sensitivities. This way, the structures of the tissues being imaged are exactly the same for all image domains, but the brightness and contrast achieved in each zone are clearly different. Yet, the relative positions of the receiving coils may be inferred from this characteristic: in the brain images, coils are located in a plane parallel to the axial view, surrounding the brain. In the spine image, the disposition of coils is longitudinal. This behavior is especially noticeable in data set PULSAR2 depicted in Fig. 3.4: the second coil yields a very poor contrast and an overall much lower brightness than the other three coils. Assuming the noise power at each image location is somehow related to the brightness in the corresponding position of the image domain, it is clear that the variability of the noise power from coil to coil for this volume can be quite large,

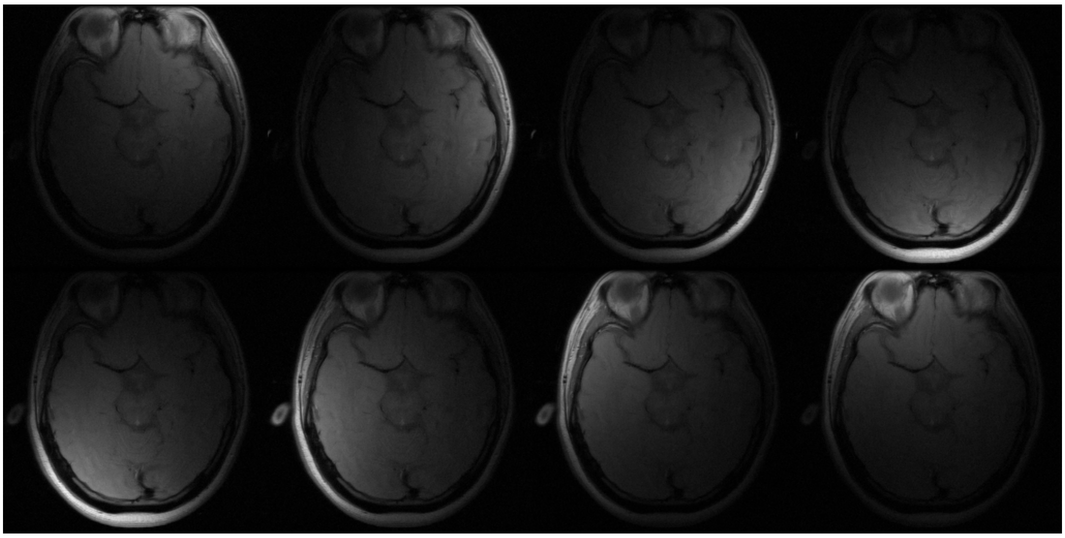


Figure 3.3: Data set PULSAR1: an axial slice in the middle brain, for each receiving coil in the 8 coils system.

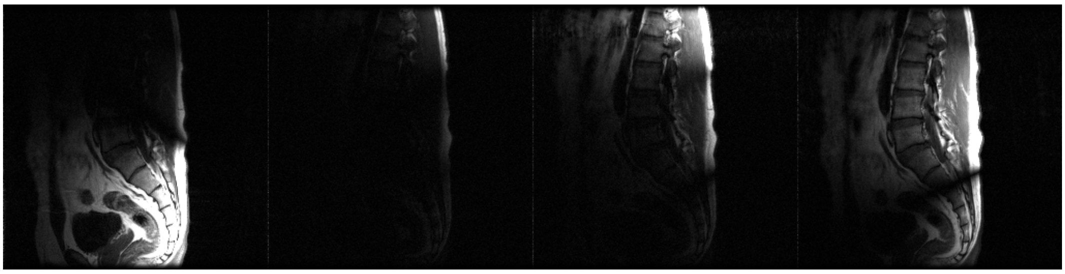


Figure 3.4: Data set PULSAR2: a slice corresponding to a human spine is shown, for each coil in the 4 coils–whole body system.

so the non–central Chi model is at stake. To formally characterize this idea, the spatial pattern of noise variances has been computed for all these data sets following the previous methodology. Instead of the noise variance itself, a normalized parameter is introduced at this point. The *stationary map* is defined as:

$$g_l^G(\mathbf{x}) = \frac{\sigma_{n_l}^2(\mathbf{x})}{\sigma_{\text{full}}^2}\sqrt{r}, \tag{3.32}$$

where σ_{full}^2 stands for the noise variance if no subsampling is performed ($\sigma_{\text{full}}^2 = \sigma_K^2/|\Omega|$), and r is the acceleration rate (downsample factor). The definition of this parameter is similar to the g–factor in SENSE, but its interpretation is not the same: $g_l^G(\mathbf{x})$ is intended to account for noise power inhomogeneities, and not for noise amplification. The reason for the normalization involving the acceleration rate is very intuitive: in case no interpolation is performed (i.e. $\omega_m(l,\mathbf{c}) = 0 \Rightarrow W_l = \Upsilon(\mathbf{x}) = 0$), the mere acceleration (subsampling) does not introduce any inhomogeneity, so $g_l^G(\mathbf{x}) = 1$. This parameter has been computed for the three GRAPPA data sets previously presented, which are $2\times$ subsampled with 13 ACS lines and GRAPPA reconstructed. Using corresponding reconstruction weights, $g_l^G(\mathbf{x})$ is computed for each coil, along with its average, its standard deviation, and its coefficient of variation (CV) across all coils. The CV is defined as:

$$C_v\{g_l^G(\mathbf{x})\} = \sqrt{\frac{\text{Var}\{g_l^G(\mathbf{x})\}}{E^2\{g_l^G(\mathbf{x})\}}} = \frac{L}{\sqrt{L-1}}\,\frac{\sqrt{\sum\limits_{l=1}^{L}\left(g_l^G(\mathbf{x}) - \frac{1}{L}\sum\limits_{m=1}^{L} g_m^G(\mathbf{x})\right)^2}}{\sum\limits_{l=1}^{L} g_l^G(\mathbf{x})}. \tag{3.33}$$

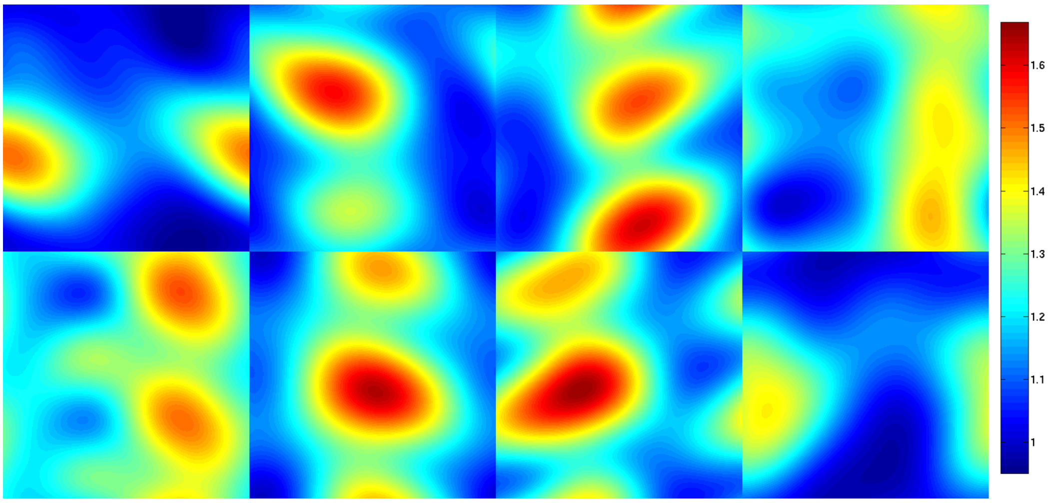


Figure 3.5: Data set BWH4: stationary map of the variance of noise for the axial slice in Fig. 3.2.

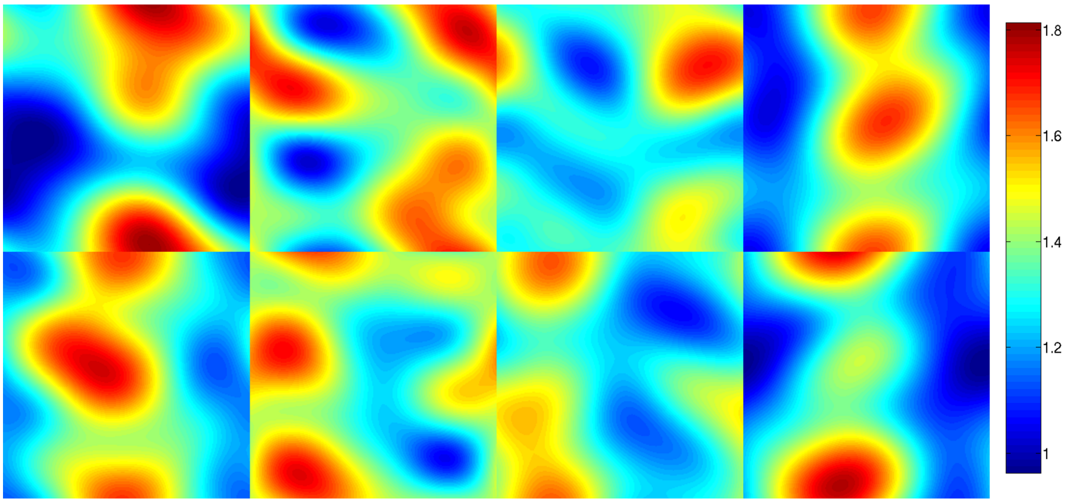


Figure 3.6: Data set PULSAR1: stationary map of the variance of noise for the axial slice in Fig. 3.3.

From the definition of $g_l^G(\mathbf{x})$ in eq. (3.32), it is trivial to show the following relations between its mean, variance, and CV and those of $\sigma_{n_l}^2(\mathbf{x})$:

$$E\{\sigma_{n_l}^2(\mathbf{x})\} = \frac{\sigma_{\text{full}}^2}{r} E\{g_l^G(\mathbf{x})\}; \tag{3.34}$$

$$\text{Var}\{\sigma_{n_l}^2(\mathbf{x})\} = \frac{\sigma_{\text{full}}^4}{r^2} \text{Var}\{g_l^G(\mathbf{x})\}; \tag{3.35}$$

$$C_v\{\sigma_{n_l}^2(\mathbf{x})\} = C_v\{g_l^G(\mathbf{x})\}, \tag{3.36}$$

so the study of the CV of $g_l^G(\mathbf{x})$ is equivalent to the study of the CV of $\sigma_{n_l}^2(\mathbf{x})$ (but the former is easier to interpret). The values of $g_l^G(\mathbf{x})$ for the GRAPPA data sets in Figs. 3.2 to 3.4 are represented in Figs. 3.5 to 3.7, respectively. In all cases, $g_l^G(\mathbf{x})$ is greater than 1 for most of the pixels, which in some sense illustrates the amplification of noise inherent to parallel reconstruction algorithms, which has been typically characterized through the g–factor in SENSE. For the brain images, its values range in $[1, 1.8]$, while the spine data shows a greater variability, ranging in $[1, 9]$. The higher variability of the spine data set is consistent with the stronger contrast changes noticed in Fig. 3.4, and justifies the previous assertion that the noise variance should be related with the brightness of the images formed by each coil at each image location. Besides, even when $g_l^G(\mathbf{x})$ (and, consequently, $\sigma_{n_l}^2(\mathbf{x})$) varies from pixel to pixel across each coil, the fluctuation is rather slow, much like what happened with the synthetic data set in Fig. 3.1: in the worst scenario, a certain local homogeneity can be assumed that assure a good behavior of neighborhood–based filters.

The other interesting study is about the variation of the noise power across the receiving coils. With this regard, the mean, variance, and CV of each data set have

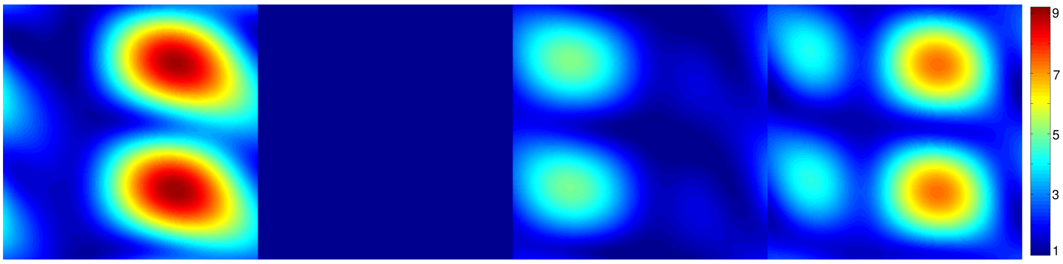


Figure 3.7: Data set PULSAR2: stationary map of the variance of noise for the slice in Fig. 3.4.

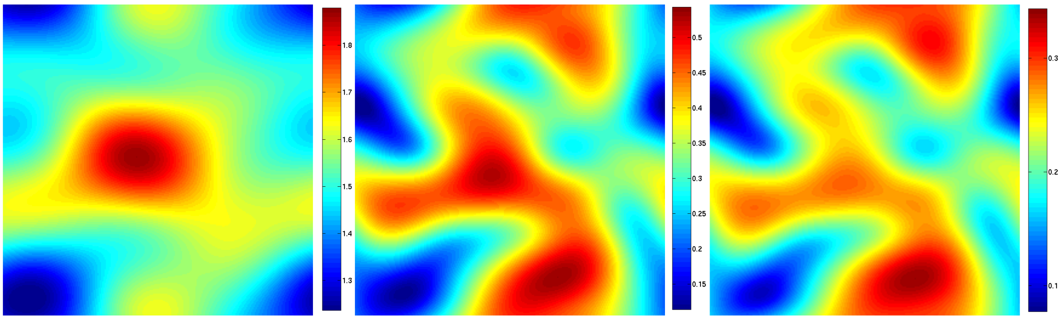


Figure 3.8: Data set BWH4: mean (left), variance (center), and coefficient of variation (right) of the stationary map of the variance of noise, for the axial slice in Fig. 3.2.

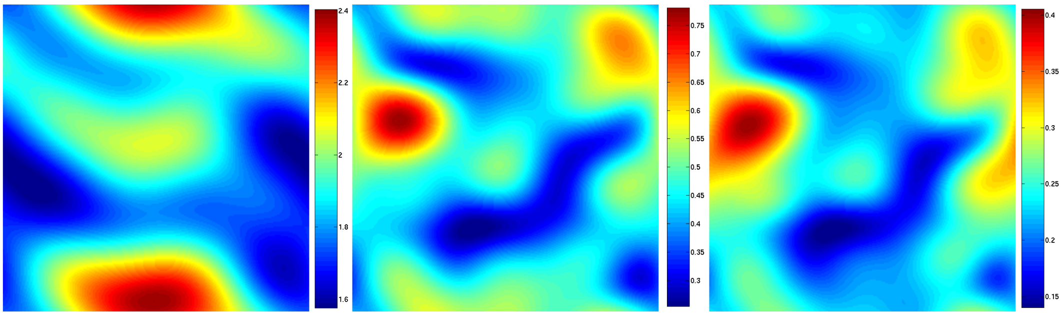


Figure 3.9: Data set PULSAR1: mean (left), variance (center), and coefficient of variation (right) of the stationary map of the variance of noise, for the axial slice in Fig. 3.3.

been computed, and represented in Figs. 3.8 to 3.10. For the CV, obviously, the closer this value to 1, the more similar the variances at each coil, and consequently the more accurate the non–central Chi model. The stationary map and its related parameters form well defined spatial patterns, whose variation is slow enough in all cases. For the brain images, the variances are always below 1, with typical values near 0.3 to 0.5. For the spine data set, the overall values of $g_l^G(\mathbf{x})$ are much higher, so both the mean and the variance are far larger for PULSAR2, with typical values in the range of tens. The most interesting parameter, however, is the CV, since it represents a variability normalized by the scale of the values being compared. The CV for PULSAR2, in fact, is more similar to the corresponding values for the other two parameters than its mean or variance are. Yet, note that the maximum of the CV for PULSAR2 is near 1.4 (see color bars), while for BWH4 and PULSAR1 it is near 0.4, i.e. it is 3.5 times smaller.

The analysis in terms of the CV is used next to characterize whether or not a non–central Chi model can be accurately fitted in each particular situation. To

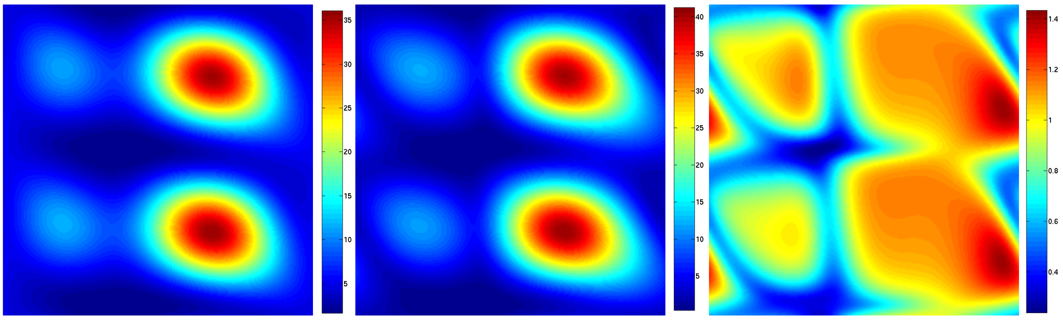


Figure 3.10: Data set PULSAR2: mean (left), variance (center), and coefficient of variation (right) of the stationary map of the variance of noise, for the slice in Fig. 3.4.

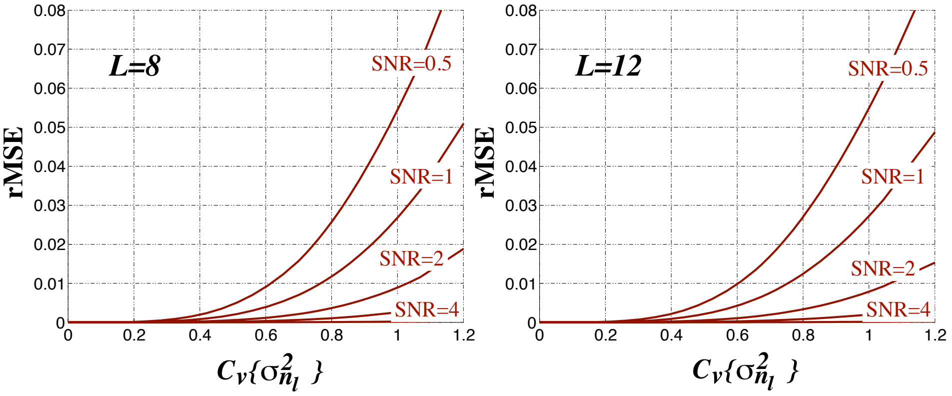


Figure 3.11: Relative errors in the PDF for the non–central Chi squared approximation, as a function of the CV of the noise power at each coil. Curves are parametrized depending on the SNR. A system with either 8 (left) or 12 (right) receiving coils is considered.

that end, the following (relative) Mean Squared Error (rMSE) is measured:

$$\text{rMSE} = \frac{\displaystyle\int_{\infty}^{\infty} |f_{X_L}(t) - \widetilde{f_{X_L}}(t)|^2 dt}{\displaystyle\int_{\infty}^{\infty} |\widetilde{f_{X_L}}(t)|^2 dt}, \tag{3.37}$$

where f_{X_L} is the true PDF of the CMS obtained by SoS (without the square root); $\widetilde{f_{X_L}}$ is an *equivalent* non–central Chi squared PDF, see eq. (3.24), where it is assumed that the variance for all coils is equal to $\overline{\sigma_n^2}$, see eq. (3.30). By virtue of Parseval's theorem, the definition in eq. (3.37) may be re–written in terms of the characteristic functions of both distributions:

$$\text{rMSE} = \frac{\displaystyle\int_{\infty}^{\infty} |G_{X_L}(w) - \widetilde{G_{X_L}}(w)|^2 dw}{\displaystyle\int_{\infty}^{\infty} |\widetilde{G_{X_L}}(w)|^2 dw}, \tag{3.38}$$

where G_{X_L} can be computed using eq. (3.26), and $\widetilde{G_{X_L}}$ using eq. (3.27). This error is calculated and plotted in Fig. 3.11 as a function of the CV of $\sigma^2_{n_l}$, for different SNR (defined as $\sqrt{A_l^2/\overline{\sigma_n^2}}$). Consider the range of CV values given by the brain data sets in the previous experiments (BWH4 and PULSAR1), i.e. $C_v \simeq 0.4$ at most. The relative errors are practically null for all SNR and coil configurations considered in Fig. 3.11: generalizing this result, it may be concluded that the non–central Chi model may be assumed for GRAPPA reconstructed images of the brain without any loss of generality. On the other hand, for higher values of the CV, the error in the approximation rapidly increases. In particular, for values ranging from 1.0 to 1.4, which are the majority in data set PULSAR2, the error is appreciable even for relatively high SNR. For 8 receiving coils, and SNR ranging in [2,4] (which may be quite realistic in DWI), the relative error can be over 1%. For high SNR, both the true PDF and the non–central Chi squared approximation tend to a Gaussian, so the error remains small.

As an additional experiment, sample data corresponding to the characteristic function in eq. (3.26) are generated. A total of 1000 experiments are carried out; for each of them, different variances $\sigma^2_{n_l}$ are chosen so that $C_v = 0.8$. The SNR is 2, and $L = 8$ coils are considered. For each experiment, sets of samples of size ranging in [1,150] are taken, and a Kolmogorov–Smirnov test is used to decide whether or not the sample data follow a non–central Chi squared distribution (with equal noise variances $\overline{\sigma_n^2}$), at a significance level $\alpha = 0.05$ [Ste74]. For comparison purposes,

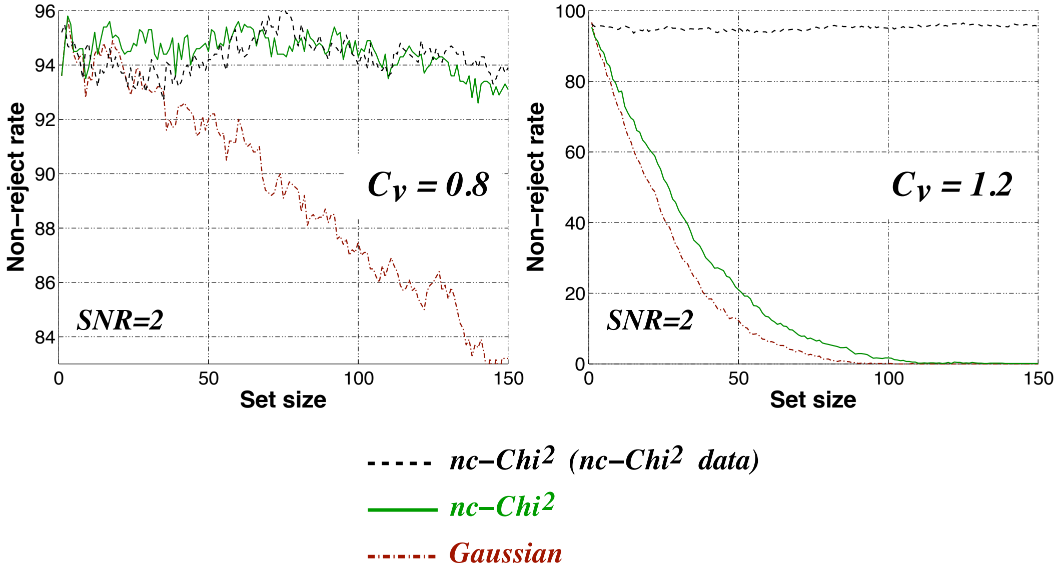


Figure 3.12: Rate of sets passing the Kolmogorov–Smirnov test for a non–central Chi squared distribution, with data generated with either different (approximated non–central Chi squared data) or the same variances (true non–central Chi squared data) at each receiving coil, as a function of the size of the population, for $C_v = 0.8$ and $C_v = 1.2$. A Gaussian distribution is also tested for comparison purposes. Significance level is $\alpha = 0.05$, and SNR$= 2$ has been used to generate data.

the test is repeated for a Gaussian distribution with mean and variance given by eqs. (3.28) and (3.29). Besides, 1000 additional experiments are set up in a similar way, generating true non–central Chi squared samples (with equal variances $\overline{\sigma_n^2}$ at each coil) and performing Kolmogorov–Smirnov tests under the hypothesis of a non–central Chi squared distribution. The whole methodology is finally repeated with $C_v = 1.2$. Fig. 3.12 represents the rate of sets (among the 1000 Montecarlo trials) for which the null hypothesis cannot be rejected (i.e. the percentage of cases for which the data can be assumed to follow the corresponding distribution). Since the significance level is $\alpha = 0.05$, the ideal value of the non–reject rate is 95%. In fact, this is approximately the value achieved for the non–central Chi squared distribution when the test is performed over true non–central Chi squared data (equal variances). For different variances with $C_v = 0.8$, the non–central Chi squared approximation is very accurate, with a very similar behavior to the ground–truth with equal variances. This assertion is justified at the sight of the curve for the Gaussian distribution: the larger the size of the sample, the easier for the test to reject the null hypothesis, so the non–reject rate of the Gaussian model rapidly decreases to values far below 95%, and it can only be accepted for very small populations. For $C_v = 1.2$, the variances at each coil are very different, and the resulting distribution is no longer similar to a non–central Chi squared. The non–reject rates for both the Gaussian and the non–central Chi squared models drop down to zero very fast, indicating that none of these models is a good approximation to the actual distribution of the sample data generated. Comparing these results with Fig. 3.11, the former scenario with SNR $= 2$ and $C_v = 0.8$ corresponds to a very small rMSE, below 0.5%, while the latter corresponds to a much higher error with rMSE $\simeq 2\%$. For the brain images presented, $C_v < 0.4$, which yields a rMSE below 0.5% for all SNR, so it can be concluded that the non–central Chi approximation is in general accurate enough for this kind of data. For the spine data set, the CV is far larger, so the validity of the model is not ensured unless voxels with large SNR are considered.

3.7 Conclusions

The main target in the present Chapter was to statistically characterize the CMS provided by the MRI scanner, since this is the signal of interest in the developments to be carried out in what follows. However, this characterization requires to study the properties of noise at each stage of the acquisition process. At the same time, this analysis has allowed to infer additional properties regarding the stationary (or nearly stationary) character of signals or the spatial correlations.

In conventional MRI, where no subsampling of the **k**–space is performed and the acquired samples can be considered independent, the noise in the complex image domain is modeled by a complex AWGN process. The power of noise is the same for all image locations. When the CMS is computed as the modulus of this process, the well–known Rician statistics arise, and the pixels at each image location are still independent from those at other locations. In multiple–coils systems without subsampling, the situation is very similar: for each coil, an entire complex image domain is reconstructed, all of them corrupted with an AWGN process with the same noise variance. When SoS is used to compute the CMS, the resulting signal follows a non–central Chi distribution, whenever correlations between different coils may be neglected (and this is usually the case).

With pMRI techniques, the final model strongly depends on the particular algorithm used. In general, those reconstructing one single image domain drive to a Rician distribution, since the CMS is computed as the modulus of a Gaussian variable. In practice, most of the existing algorithms, except for GRAPPA, provide a unique **x**–space, so Rician statistics are still of great interest. Contrary to conventional MRI, the noise in the image cannot be considered stationary strictly speaking, since its variance shows fluctuations across the image. Besides, noisy samples cannot be considered independent either, due to interpolation artifacts.

The analysis of GRAPPA reconstructed signals is more difficult. Although its feature of providing one complete image domain for each receiving coil assimilates it to conventional multiple coils systems, the non–central Chi distribution does not directly apply to this case. Apart from the aforementioned interpolation artifacts related to non–stationary noisy patterns and correlations, the noise variance for each receiving coil is different with this kind of reconstruction. Hence, a non–central Chi model cannot be assumed in general. For brain images, which are the data sets of interest in this dissertation, it has been shown that in fact this distribution is a very accurate approximation to the actual statistics of noise. For other kinds of data, such as the spine of PULSAR2, this is not the case. The characterization of GRAPPA signals has allowed also to study the spatial variability of the the statistics of noise (variance) across the MRI volumes. It has been shown that the variability of the noise power is not very important, and yet it is slow enough to allow the use of image processing algorithms assuming stationary signals. Other pMRI algorithms can be mapped to reconstructions in the **k**–space (for example, the mSENSE extension to SENSE), so this same property is expected to be shared by all of them: although they have to be described by Rician statistics, since the computation of the CMS is based on a unique **x**–space, the interesting part is that a quasi–stationary model can always be assumed.

The Chapter has been primarily focused on multiple coils systems and GRAPPA, mainly because the theoretical foundations for the latter have not been thoroughly studied before. In Diffusion Imaging, multiple coils systems without subsampling are not as interesting as pMRI algorithms, due to the necessity to accelerate the acquisition process. Among these techniques, GRAPPA (together with SENSE) is the most popular due to its penetration in commercial machinery. Hence, the importance of studying non–central Chi statistics, and generalizing existing image processing methods to this kind of signals, is fully justified.

It is convenient to stress the limitations of the study performed in this Chapter. As mentioned above, the main ones are those related to correlations between the signals recovered by each coil in the scanner. This problem is even more important with GRAPPA, since the interpolation of the **k**–space introduces additional dependencies of data. Besides, the analysis and validation have been carried out taking into account anatomical MRI data sets. With diffusion data sets, the difference relies on the inclusion of additional pulsed gradients in the sequences; regardless on the particular sequences described by the protocol being used, values in the **k**–space are always acquired as samples of the phase/frequency encoded signal received by a RF antenna, so no fundamental differences should be present. Finally, certain MRI techniques, such as EPI, which can show their own peculiarities with regard to the statistical characterization of noise, has been removed from the study.

4

Impact of noise in Diffusion Imaging

The previous Chapter was focused on the analysis of the statistics of noise in the complex $\mathbf{x}$–space and the CMS. In this Chapter, it is analyzed how this noise is propagated from the $\mathbf{q}$–space (at each point of the $\mathbf{x}$–space) to its corresponding $\mathbf{R}$–space. It is demonstrated that the particular statistics of noise in the CMS entail an important bias in the estimation of the diffusion propagator. This bias cannot be eliminated after the estimation, so it has to be accounted for at the level of DWI (i.e., in the $\mathbf{x} \times \mathbf{q}$–space). The analysis is carried out through the case study of DTI with diffusion tensor estimation based on Weighted Least Squares (WLS), and then extrapolated to more general, non–Gaussian estimators of the diffusion propagator. In particular, it is shown that the bias in the estimation of the components of the diffusion tensor may be as important as its variance for non–central Chi signals. This implies that the use of more gradient directions, up from a given number, may be useless to reduce the error in the estimation, stressing the importance of the unbiased filtering techniques proposed in Chapter 5. This case study corresponds to one of the most popular Difussion Imaging techniques nowadays, so it constitutes an interesting analysis even beyond the scope of this dissertation. The original contribution in this Chapter is threefold: first, the theoretical prediction of the impact of Rician noise in the estimation of the diffusion tensor, which until now had been described only in terms of empirical results. Second, the generalization of this study to multiple–coils/pMRI systems. Third, the guidelines to decide whether it is useful to increase the number of gradient directions or the number of repetitions (NEX) in a given situation.

4.1 Introduction

Rician distribution is an accurate model for the CMS in non–parallel, single coil MRI. As described in the previous Chapter, this same model holds for most of the existing pMRI techniques (SENSE, SMASH), whenever a unique complex $\mathbf{x}$–space is recovered for all receiving coils, either from reconstructions on the $\mathbf{k}$–space, the $\mathbf{x}$–space, or both. However, for non–parallel, multiple coil MRI, a whole $\mathbf{x}$–space is available for each coil, and the CMS is computed as the combination of all of them. For parallel reconstructions based on GRAPPA, a whole $\mathbf{k}$–space is reconstructed for each coil, so the situation is very similar. In these cases, the SoS yields a non–central Chi distribution for the CMS. Besides, the Rician model is a particular case of the non–central Chi, so the latter may be seen as a general model for the most common parallel and non parallel acquisition schemes.

One known issue with these distributions is their bias. The expected value of the distribution is not the actual value of the noise–free CMS, and as a consequence the measurement of the corrupted CMS introduces a systematic error in the estimation of the true CMS. The effect of this bias in Diffusion Imaging has been previously reported in the literature for the case of Rician distributed

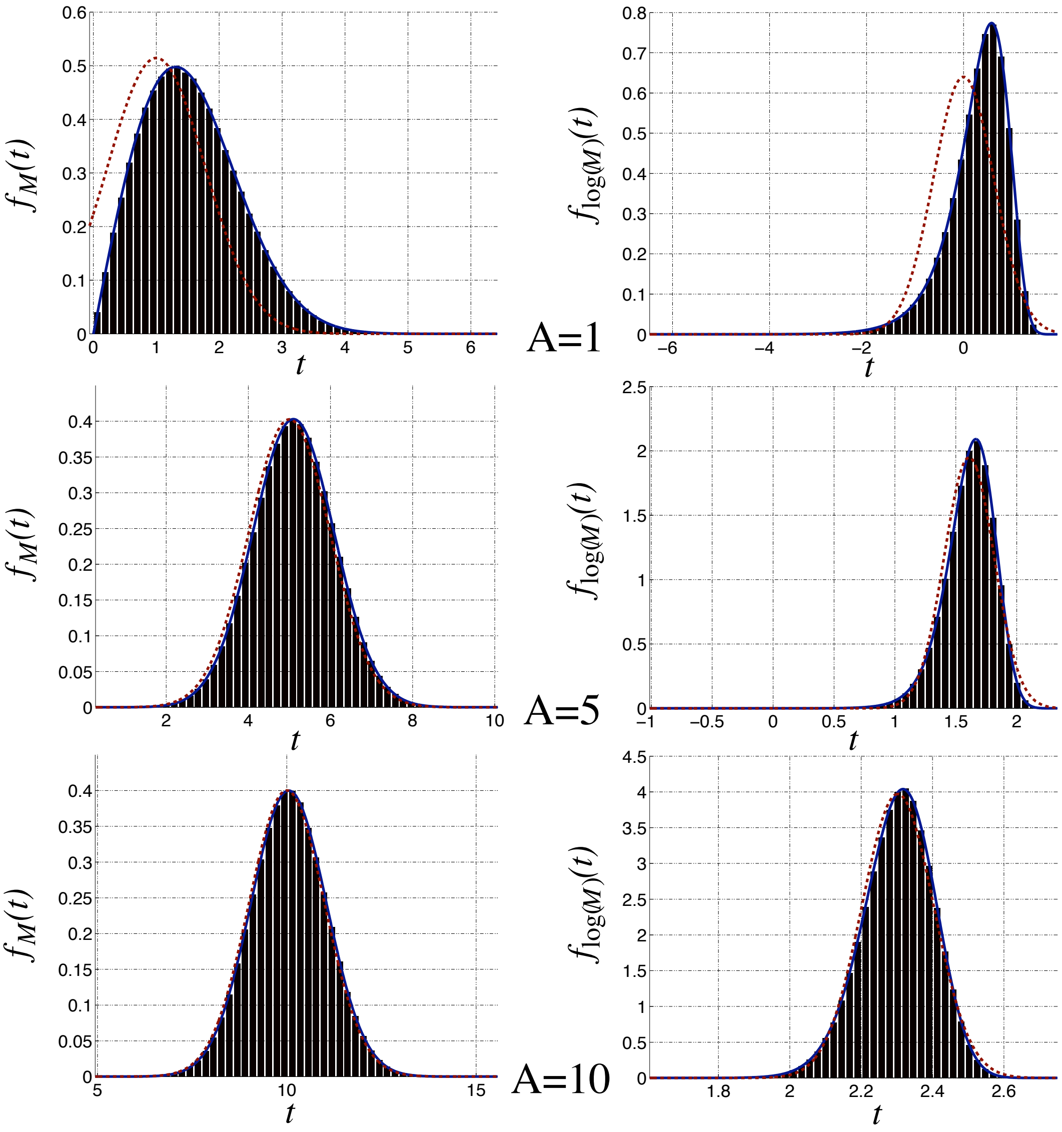


Figure 4.1: The PDF of Rician (left) and log–Rician (right) data for different values of A and $\sigma_n = 1$ (thus, for different SNR). Blue straight lines represent the theoretical PDF, while the bar plots have been drawn from 10^6 Montecarlo trials in each case. Red discontinuous lines represent a Gaussian PDF with the same mean and variance as the corresponding Rician or log–Rician PDF. It may be seen that both distributions tend to a Gaussian distribution very fast.

signals [Bas00a, Jon04, Sal05]. These studies are merely empiric and focused on DTI, but they have shown that Rician noise introduce a systematic error (underestimation) in the computation of the FA which cannot be removed after the diffusion tensor has been estimated. Similar results have been reported in [Bas06] for the MD and the FA. Although these conclusions are grounded in empirical studies, they all suggest that noise in the CMS is propagated from the **q**–space to the **R**–space rendering errors which cannot be recovered.

The most common technique to estimate the diffusion tensor in DTI is Least Squares (LS) or, more often, Weighted Least Squares (WLS) [Sal05], based on the formulation in eq. (2.32). In this case the logarithm of the DWI is computed, so the signals no longer follow Rician or non–central Chi distributions, but log–Rician or log–non central Chi. It has been shown that, although the bias of original DWI persists in the log–compressed images, log–Rician distributed signals are nearly unbiased for reasonably high SNR [Sal05] (see Fig. 4.1). It follows that the artifacts associated to the bias in the CMS for DTI, at least for Rician distributed noise, are relatively subtle.

For non–Gaussian diffusion propagators (see Section 2.7), it is common to work directly with the CMS and not with its logarithm. It has been previously reported in [Cla08] that, in this case, the Rician noise propagated from the **q**–space to the **R**–space noticeably distorts the computation of orientation information, even for moderate–high SNR, enclosing an undesired angular blurring.

On the other hand, although the effects of Rician noise have been empirically assessed, the impact of non–central Chi noise in Diffusion Imaging is not clear. Moreover, all the aforementioned studies are empirical, and no theoretical analysis has been performed to yield solid conclusions on this issue. In this Chapter, the aim is, first, to carry out an analytic study on the propagation of noise for the case study of Gaussian diffusion, i.e. for DTI. Second, the study is generalized to non–central Chi signals, and it is stated that the bias is even more important for this statistics. The case study is centered on DTI based on WLS estimation of the diffusion tensor; this is without any doubt the most widespread Diffusion Imaging technique nowadays, so in fact this survey constitutes itself an interesting result. For more general HARDI techniques, an analytic study is not feasible for two reasons:

- The estimation of the diffusion propagator from the signal in the **q**–space is in general highly non–linear or even iterative.
- The techniques reviewed in Section 2.7 have very different foundations and characteristics, so a general approach gathering all of them is not possible.

Nonetheless, the work in [Cla08] evidences that the effect of noise for this kind of estimators is more important than it is for DTI. This result may be used to infer the impact of non–central Chi (and thus of Rician) distributed signals in HARDI estimators of the diffusion propagator. In particular, the importance of correctly accounting for noise effects in the estimators presented in Chapters 6 and 7 will be fully justified.

4.2 Characterization of log–non–central Chi signals

4.2.1 Closed forms for the mean and variance

As described in Chapter 3, the statistics of the CMS when SoS is used follow a non–Central Chi distribution, whose PDF is reproduced here for convenience:

$$f_{M_L}(t;A_L,\sigma_n,L) = \frac{A_L}{\sigma_n^2}\left(\frac{t}{A_L}\right)^L \exp\left(-\frac{t^2+A_L^2}{2\sigma_n^2}\right) I_{L-1}\left(\frac{A_L\, t}{\sigma_n^2}\right) u(t), \tag{4.1}$$

where M_L is the measured CMS; σ_n^2 is the noise power at each point $\mathbf{x}$ of the complex $\mathbf{x}$–space (it is assumed to be roughly the same for all receiving coils); L is the number of receiving coils; I_{L-1} is the modified Bessel function of the first kind an order $L-1$; $u(t)$ is Heaviside's step function; $A_L = \sqrt{\sum_{l=1}^{L}(A_{l_c}^2 + A_{l_s}^2)}$ is the ideal value of the CMS reconstructed from the L signals at each coil by SoS (see eq. (3.21)). For $L=1$, eq. (4.1) reduces to the Rician distribution. The mean and variance of this distribution may be computed in closed form [Die08]:

$$\begin{aligned} E\{M_L\} &= \sqrt{2}\sigma_n \frac{\Gamma(L+1/2)}{\Gamma(L)}\,{}_1F_1\left(-\frac{1}{2},L,-\frac{A_L^2}{2\sigma_n^2}\right); && (4.2)\\ \mathrm{Var}\{M_L\} &= A_L^2+2L\sigma_n^2-E\{M_L\}^2, && (4.3) \end{aligned}$$

where Γ stands for Euler's Gamma function, and ${}_1F_1$ is the confluent hypergeometric function of the first kind [Abr72]. As stated in Section 4.1, the statistics of interest are those of $\log(M_L)$ and not M_L. The corresponding PDF may be computed from eq. (4.1) using the fundamental theorem of probability:

$$\begin{aligned} f_{\log M_L}(t;A_L,\sigma_n,L) &= f_{M_L}(\log^{-1}t;A_L,\sigma_n,L)\frac{d\log^{-1}t}{dt}\\ &= \frac{A_L^2}{\sigma_n^2}\left(\frac{e^t}{A_L}\right)^{L+1}\exp\left(-\frac{e^{2t}+A_L^2}{2\sigma_n^2}\right)I_{L-1}\left(\frac{A_L\,e^t}{\sigma_n^2}\right). && (4.4) \end{aligned}$$

The corresponding mean and variance may be computed in the following form (see appendix A):

$$\begin{aligned} E\{\log(M_L)\} &= \frac{1}{2}\log\left(2\sigma_n^2\right)+\frac{1}{2}\frac{A_L^2}{2L\sigma_n^2}\,{}_2F_2\left(1,1:2,1+L;-\frac{A_L^2}{2\sigma_n^2}\right)+\frac{1}{2}\psi(L); && (4.5)\\ \mathrm{Var}\{\log(M_L)\} &= \frac{1}{4}\left(\widetilde{N}_L\left(\frac{A_L^2}{2\sigma_n^2}\right)-2\log\left(2\sigma_n^2\right)\frac{A_L^2}{2L\sigma_n^2}\,{}_2F_2\left(1,1:2,1+L;-\frac{A_L^2}{2\sigma_n^2}\right)\right.\\ &\quad\left.-\left(\psi(L)+\frac{A_L^2}{2L\sigma_n^2}\,{}_2F_2\left(1,1:2,1+L;-\frac{A_L^2}{2\sigma_n^2}\right)\right)^2\right); && (4.6) \end{aligned}$$

with:

$$\widetilde{N}_L(z)=e^{-z}\sum_{k=0}^{\infty}\frac{z^k}{k!}\left(\psi(k+L)^2+\psi^{1)}(k+L)\right),$$

where $\psi(x)$ is the polygamma function, and $\psi^{1)}(x)$ is its first derivative. Eq. (4.5) generalizes the expression given in [Sal05, eq. (11)] for the mean of the log–Rician distribution, while eq. (4.6) has not been previously reported. Indeed, the previous expressions can be particularized to $L=1$ (see appendix A) to obtain:

$$\begin{aligned} E\{\log(M)\} &= \frac{1}{2}\Gamma^{(l)}\left(0,\frac{A^2}{2\sigma_n^2}\right)+\log(A); && (4.7)\\ \mathrm{Var}\{\log(M)\} &= \frac{1}{4}\left(\widetilde{N}_1\left(\frac{A^2}{2\sigma_n^2}\right)+\left(\Gamma^{(l)}\left(0,\frac{A^2}{2\sigma_n^2}\right)+\log\left(\frac{A^2}{2\sigma_n^2}\right)\right)^2\right). && (4.8) \end{aligned}$$

The expected value of $\log(M_L)$ tends to $\log(A_L)$ for high SNR, as depicted in Fig. 4.1 for the log–Rician case. This is the reason why, for high SNR, the estimation of the diffusion tensor from Rician distributed DWI is nearly unbiased.

4.2.2 Operative linear approximations

The complexity of eqs. (4.5) and (4.6) does not allow a detailed analysis, so practical approximations have to be developed [TV09c]. In what follows, it is assumed that the CMS for multiple receiving coils has the form:

$$A_L=\sqrt{\sum_{l=1}^{L}A_{l_c}^2+A_{l_s}^2}\equiv\sqrt{\sum_{l=1}^{L}A^2+0^2}=\sqrt{L}\,A, \qquad (4.9)$$

i.e. the same real signal A is ideally acquired by all receiving coils l: $A_{l_c}=A$, $A_{l_s}=0$. This assumption does not impose any loss of generality, but the notation in the developments to be carried out is far simpler. The Taylor series expansion of $\log(M_L)$ for $\sigma_n \ll A_L$ is:

$$\begin{aligned}\log\left(M_L\right) &= \log\left(A_L\right) + \frac{1}{AL}\sum_{l=1}^{L} n_{l_c} + \frac{1}{2A^2L^2}\sum_{k=1}^{L}\sum_{l=1}^{L}\left(L\delta_{kl}-2\right)n_{k_c}n_{l_c}\\ &+ \frac{1}{2LA^2}\sum_{l=1}^{L} n_{l_s}^2 + \mathcal{O}\left(\frac{\sigma_n^3}{A^3}\right),\end{aligned} \tag{4.10}$$

with δ the Kronecker delta function and n_{l_c}, n_{l_s} the phase and quadrature (respectively) zero–mean, Gaussian noises corresponding to the real and imaginary parts of the complex signal in the $\mathbf{x}$–space. From eq. (4.10), it follows:

$$E\left\{\log\left(M_L\right)\right\} = \log\left(A_L\right) + (L-1)\frac{\sigma_n^2}{A_L^2} + \mathcal{O}\left(\frac{\sigma_n^3}{A^3}\right). \tag{4.11}$$

In fact, since the expectation of all terms in $\mathcal{O}(\sigma_n^3/A^3)$ represent odd–order moments of Gaussians, $\mathcal{O}(\sigma_n^4/A^4)$ could have been written instead. The MSE in the estimation may be computed as the variance plus the squared bias, so the expansion of the variance to order 4 (comparable to bias$^2 = \mathcal{O}(\sigma_n^4/A_L^4)$) has to be computed. This requires to compute the expansion of $\log(M_L)$ to order 3. After some calculations, it yields:

$$\mathrm{Var}\left\{\log\left(M_L\right)\right\} = \frac{1}{2}a^{-1} - \frac{3L-4}{4}a^{-2} + \mathcal{O}\left(a^{-3}\right); \tag{4.12}$$

$$\mathrm{bias}^2\left\{\log\left(M_L\right)\right\} = \frac{(L-1)^2}{4}a^{-2} + \mathcal{O}\left(a^{-3}\right), \tag{4.13}$$

where the variable:

$$a = \frac{A_L^2}{2\sigma_n^2} \tag{4.14}$$

is a normalized SNR (according to the standard notation, $a = \mathrm{SNR}^2/2$). For the Rician case ($L = 1$), the squared bias is in the order of $\mathcal{O}(a^{-3}) = \mathcal{O}(\sigma_n^6/A_L^6)$, so it is not so important, as it has been noted in [Sal05] (see Fig. 4.1). For $L > 1$, the bias may be important if L is in the same order of magnitude as A_L/σ_n.

4.3 Tensor fitting based on Weighted Least Squares

As stated in Chapter 2, the Stejskal–Tanner model [Ste65] may be used to linearize the estimation of the diffusion propagator. Taking into account the noisy nature of the CMS, eq. (2.30) may be re–written in the form:

$$\begin{aligned}\log\left(A_{L,0}\right) - \log\left(M_{L,i}\right) &= \left[g_{i,1}^2,\ 2g_{i,1}g_{i,2},\ 2g_{i,1}g_{i,3},\ g_{i,2}^2,\ 2g_{i,2}g_{i,3},\ g_{i,3}^2\right]\\ &\cdot\left[b\mathscr{D}_{11},\ b\mathscr{D}_{12},\ b\mathscr{D}_{13}, b\mathscr{D}_{22},\ b\mathscr{D}_{23},\ b\mathscr{D}_{33}\right]^T + \varepsilon_i,\end{aligned} \tag{4.15}$$

for $i = 1\ldots N$ with N the number of diffusion gradient directions. The noise ε is considered additive (which is roughly true for high SNR as suggested in Fig. 4.1). Note that the complex signal $S(\mathbf{g})/S(\mathbf{0})$ may be substituted by its modulus by virtue of the result in eq. (2.27), so the CMS may be used instead. In eq. (4.15), the subindex i is used in A_L and M_L to denote the gradient direction they correspond to (0 corresponds to the non–weighted baseline). As opposed to [Sal05], $A_{L,0}$ is not included in the estimation, i.e., it is considered that the impact of noise in the baseline $A_{L,0}$ is negligible ($M_{L,0} \simeq A_{L,0}$). The baseline is generally far less noisy than the gradient images, and moreover, several baseline images are often available.

Hence, the error due to $A_{L,0}$ is negligible compared to the error due to $A_{L,i}$, and it does not appreciably affect the analysis. Nevertheless, a similar study to that here presented may be carried out including $A_{L,0}$ in the estimation, with almost identical conclusions. Given the problem statement in eq. (4.15) as an overdetermined (if $N > 6$) set of linear equations with noisy observations, the natural solution is the use of LS, or more precisely WLS. WLS is useful since the variance of each observation $M_{L,i}$ is different, as remains evident from eq. (4.12). Its formulation is as follows:

$$G^T W (\mathbf{Y} - G\mathbf{X}) = 0 \Rightarrow \mathbf{X} = \left(G^T W G\right)^{-1} G^T W \mathbf{Y}, \tag{4.16}$$

where $\mathbf{Y}$ is an $N \times 1$ vector representing each of the DWI data $\log(M_{L,i}) - \log(A_{L,0})$, $\mathbf{X}$ is the 6×1 vector of unknowns ($b\mathscr{D}_{ij}$), G is the $N \times 6$ matrix resulting from the concatenation of each row in eq. (4.15), $[g_{i,1}^2, \ldots, g_{i,3}^2]$, and W is the $N \times N$ matrix of weights. The Gauss–Markov theorem states that, under very weak assumptions, the WLS solution is the minimum variance estimate if W is chosen to be the inverse of the covariance matrix of data, $C_{\mathbf{YY}}$. If the estimator is unbiased (as is approximately the case for Rician noise), WLS is in fact the Best Linear Unbiased Estimator (BLUE). Since the noise in each gradient image is assumed to be independent, $C_{\mathbf{YY}}$ reduces to a diagonal matrix with diagonal elements equal to the variance of each term $\log(M_{L,i}) - \log(A_{L,0})$. This variance is the same as that of $\log(M_{L,i})$, since the noise in $A_{L,0}$ has been neglected. Therefore, $W = \text{diag}(W_{ii})$, and from eq. (4.13):

$$W_{ii} = \text{Var}^{-1}\{\mathbf{Y}_i\} \simeq \frac{1}{a_i^{-1}/2 - a_i^{-2}(3L-4)/4} \simeq 2a_i + (3L-4). \tag{4.17}$$

In what follows, the terms of order a_i^{-2} will be neglected, so that $W_{ii} = 2a_i = A_{L,i}^2/\sigma_n^2$. There are two reasons for this:

1. This formulation is identical to the traditional WLS for Rician noise, see [Sal05], and is the one implemented in all software tools for the estimation of the diffusion tensor. Note that even in the case of Rician noise, the term in a^{-2} is not null (but very small), and even so WLS is always performed with this formulation.

2. The weights W_{ii} have in fact to be only proportional to $2a_i$. With this formulation, it is not necessary to know the exact value of σ_n, since any weight proportional to $A_{L,i}^2$ may be used; if the term in a^{-2} is included, it is necessary to estimate σ_n at each image location and to include the number of coils, L, as an additional parameter of the algorithm.

Once again, the whole analysis may be performed with the exact expression for W_{ii} with very similar conclusions. Finally, note that the weights W_{ii} depend on $A_{L,i}$ (not $M_{L,i}$), so they cannot be known *a priori*; this pitfall may be obviated by iteratively estimating $\mathbf{X}$ and updating the value of W [Sal05]. In the remaining it is assumed that W_{ii} are known without uncertainty.

4.4 Bias and variance of the tensor components

Once the model for the estimation of the diffusion propagator has been presented (see Section 4.3), it remains to study the propagation of noise from the **q**–space to the **R**–space. Let $\mathscr{L} = (G^T W G)^{-1}$; from eqs. (4.11), (4.12), (4.16), and (4.17), the covariance matrix of $\mathbf{X}$ reads:

$$\begin{aligned}
C_{\mathbf{XX}} &= E\left\{(\mathbf{X}-E\{\mathbf{X}\})(\mathbf{X}-E\{\mathbf{X}\})^T\right\} = \mathscr{L}G^TWC_{\mathbf{YY}}W^TG\mathscr{L}^T \\
&= \mathscr{L}G^TW\mathrm{diag}\left(\frac{1}{2}a_i^{-1}-\frac{\Theta(L)}{4}a_i^{-2}\right)WG\mathscr{L} \\
&= \mathscr{L}G^TWG\mathscr{L}-\frac{\Theta(L)}{4}\mathscr{L}G^TW\mathrm{diag}\left(a_i^{-2}\right)WG\mathscr{L} \\
&= \mathscr{L}-\Theta(L)\mathscr{L}G^TG\mathscr{L}, \qquad (4.18)
\end{aligned}$$

where $\Theta(L)=3L-4$ and W, $\mathscr{L}$ are symmetric. Note that the second term in eq. (4.18) appears due to the fact that $W_{ii}=2a_i$ is used instead of the *true* value of $\mathrm{Var}^{-1}\{Y_i\}$, and it is not null even for the Rician case[1]. For the bias, since $\mathrm{bias}\{\mathbf{Y}_i\}=a_i^{-1}(L-1)/2=(L-1)/W_{ii}$:

$$\mathrm{bias}\{\mathbf{X}\}=(G^TWG)^{-1}G^TW\mathrm{bias}\{\mathbf{Y}\}=(L-1)\mathscr{L}G^T\mathbf{1}, \qquad (4.19)$$

where $\mathbf{1}$ is an $N\times 1$ vector of ones. The 6×1 vector $G^T\mathbf{e}_1$ may be easily characterized if the gradient directions are assumed to be uniformly distributed on the sphere. Each component $g_k,\ k=1,2,3$ may be seen as a uniform random variable, so:

$$\left.\begin{array}{r}
g_k\sim U(-1,1) \\
\sum_{i=1}^{N} g_{i,k}^2\simeq NE\{g_k^2\}=N/3 \\
\sum_{i=1}^{N} 2g_{i,j}g_{i,k}\simeq 2NE\{g_j\}E\{g_k\}=0
\end{array}\right\}\Rightarrow G^T\mathbf{1}=\frac{N}{3}[1,0,0,1,0,1]^T=N\mathbf{v}, \qquad (4.20)$$

for $\mathbf{v}=[1/3,0,0,1/3,0,1/3]^T$. In fact, this assumption is very realistic for gradient directions chosen as antipodal pairs (the most common case), and it may be empirically shown that eq. (4.20) describes a real situation pretty accurately. Finally, the MSE is defined as:

$$\begin{aligned}
\mathrm{MSE} &= \mathrm{Var}\{\mathbf{X}\}+\mathrm{bias}^2\{\mathbf{X}\}=b^2\left(\Delta_{11}^2+\Delta_{12}^2+\Delta_{13}^2+\Delta_{22}^2+\Delta_{23}^2+\Delta_{33}^2\right) \\
&= \mathrm{tr}(\mathscr{L})-\Theta(L)\mathrm{tr}\left(\mathscr{L}G^TG\mathscr{L}\right)+(L-1)^2N^2\mathbf{v}^T\mathscr{L}^2\mathbf{v}, \qquad (4.21)
\end{aligned}$$

for $\Delta_{ij}=\mathscr{D}_{ij}-\widetilde{\mathscr{D}}_{ij}$ and $\widetilde{\mathscr{D}}_{ij}$ the WLS estimate of $\mathscr{D}_{ij}$.

4.5 Numerical results

4.5.1 Bias and variance in the DWI components

In this Section, the accuracy of the approximations given by eqs. (4.12) and (4.13) for the noise in the $\mathbf{q}$–space is tested. In Fig. 4.2 (left), the true values given by eqs. (4.5) and (4.6) are represented, together with empirically obtained values (for 2000 independent samples for each a and L). The proposed approximations are depicted superimposed to show their high accuracy, except for very low (unrealistic) values of a. The larger L, the larger the value of a below which the approximation diverges from the actual value. Nevertheless, note that the difference is only noticeable for $L=8$ (below $a=60$). As shown later on (and in Fig. 4.2, right) for $L=8$ the values of interest for a are in the order of 140; for larger values of L, the approximations are inexact for higher a, but at the same time the study of the bias is also of interest for higher a, so the approximation is extremely accurate in all cases.

---

[1] Yet, $\Theta(L)$ is negative for $L=1$, so for the Rician case the variance is *greater* than $\mathscr{L}$, contrary to the case $L>1$.

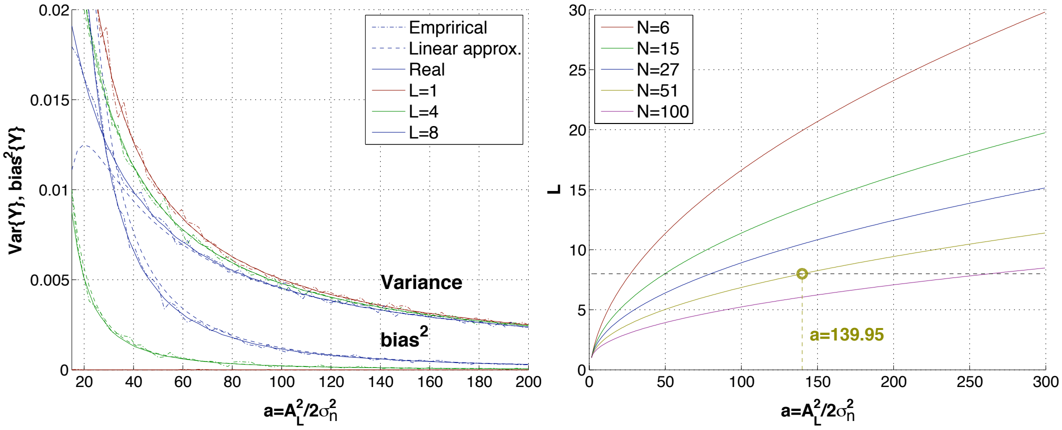


Figure 4.2: Left: bias and variance in the DWI signals, as a function of a for different numbers of coils; true and empirically computed values, together with linear approximations, are represented. **Right**: minimum number of receiving coils required (for each a and N) so that the (squared) bias equals the variance in the diffusion tensor components (see eq. (4.24)).

4.5.2 Bias and variance in isotropic tensor components

The foregoing presents an analysis on the effect of the bias of the non–central Chi distribution in the computation of the diffusion tensor. The aim is to show that the bias in the **q**–space propagates to the **R**–space; yet, it is even amplified rendering important errors in the estimation even for high SNR. First, a simplified scenario of isotropic diffusion, where all diffusion directions behave the same, is considered:

$$W_{ii} = 2a_i = 2a = \text{SNR}^2. \tag{4.22}$$

The MSE reduces in this case to:

$$\text{MSE} = \left(\frac{a^{-1}}{2} - \frac{a^{-2}}{4}\Theta(L)\right) \text{tr}\left(\left(G^TG\right)^{-1}\right) + a^{-2}\frac{(L-1)^2N^2}{4} v^T \left(G^TG\right)^{-2} v, \tag{4.23}$$

since $\mathscr{L} = a^{-1}/2(G^TG)^{-1}$. The behavior of G^TG has to be analyzed to study the relation between the bias and the variance. A similar reasoning to that made for the calculation of v in eq. (4.20) shows that it always has the eigenvector $v/\|v\|$, associated to the eigenvalue $N/3$, for antipodal pairs. Besides, the approximation: $\text{tr}((G^TG)^{-1}) \simeq 29.3/N$ is extremely accurate[2] for all values of N. With these equivalences, the MSE may be re–written:

$$\begin{aligned} \text{MSE} &\simeq \frac{29.3}{N}\left(\frac{a^{-1}}{2} - \frac{a^{-2}}{4}(3L-4)\right) + a^{-2}\frac{9(L-1)^2N^2}{4N^2} v^T \frac{vv^T}{\|v\|^2} v \\ &= \frac{29.3}{N}\left(\frac{a^{-1}}{2} - \frac{a^{-2}}{4}(3L-4)\right) + a^{-2}\frac{3(L-1)^2}{4}, \end{aligned} \tag{4.24}$$

since $(G^TG)^{-2}$ has the eigenvector $v/\|v\|$ associated to $9/N^2$, and all other eigenvectors are orthogonal to v. While the variance diminishes with the number of gradients N, this is not the case for the bias: Fig. 4.2 (right) shows the number of coils needed for each a and N so that the term corresponding to the bias in eq. (4.24) equals the term corresponding to the variance. For example, with $L = 8$ coils and 51 gradient directions, it is shown that, if $a < 139.95$, the bias is more important than the variance. Alternatively, for $a < 139.95$ and $L = 8$, it makes little sense to take more than 51 gradients, since the variance is reduced but not the

---

[2] This estimate has been carried out empirically, since no theoretical foundation has been found for it. However, note that it does not affect the general validity of the analysis performed; yet, it is not necessary –or useful– for the study presented in Section 4.5.3.

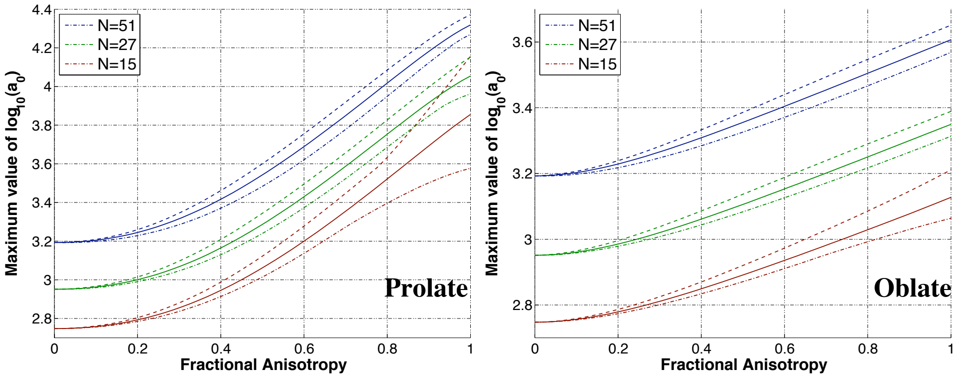


Figure 4.3: Log–plot of the maximum value of $a_0 = A_{L,0}^2/2\sigma_n^2$ which makes the squared bias equal to the variance of the components of the diffusion tensor, for different tensor shapes. Minimum, mean, and maximum values among all possible orientations are shown.

bias, so the acquisition time is increased unfruitfully. For $a = 50$, $L = 8$, the situation is the same for more than 15 gradients. It follows that in some situations it may be more convenient to increase the number of repetitions (NEX) in the acquisition than using more gradient directions. This has the effect of dividing the noise power σ_n^2 by NEX (since the signals are averaged in the **k**–space, where the noise is Gaussian distributed, see Chapter 3), so that:

$$\text{MSE} = \frac{29.3}{N \cdot \text{NEX}} \left(\frac{a^{-1}}{2} - \frac{a^{-2}}{4\ \text{NEX}}(3L-4) \right) + a^{-2} \frac{3(L-1)^2}{4\ \text{NEX}^2}. \tag{4.25}$$

A constant value of $N \cdot \text{NEX}$ takes a constant acquisition time, and besides provides a constant reduction factor for the variance, but the bias depends only on NEX (squared). It may be concluded that, for a given combination of a, N and L which makes the squared bias similar to the variance (see Fig. 4.2), it is preferable to increase NEX than using more gradient directions.

4.5.3 Generalization to anisotropic tensors

When the weigths W_{ii} are not equal (the usual anisotropic situation), the analysis is more complex, since it directly depends on all the involved variables (L, N, the b–value, and the tensor components themselves) in a non trivial way. An illustrative experiment with typical values $L = 8$; $N = 15$, 27, and 51; $b = 1500\text{s/mm}^2$; $\text{MD} = 0.8 \cdot 10^{-3}\text{mm}^2/\text{s}$ has been intended. The maximum value of a_0 (defined as $a_0 = A_{L,0}^2/2\sigma_n^2$, with $A_{L,0}$ the amplitude of the baseline image) for which the contribution of the bias is greater or equal than the contribution of the variance is computed using eq. (4.21). Since this result depends on the orientation of the diffusion tensor, Fig. 4.3 shows maximum, minimum, and mean values of a_0 for different FA. Two cases are considered: a prolate tensor ($\lambda_1 > \lambda_2 = \lambda_3$) and an oblate tensor ($\lambda_1 = \lambda_2 > \lambda_3$). For FA=0 (isotropic diffusion, equal weights), the situation is the same as in Fig. 4.2 (right). With non–isotropic tensors, the bias becomes even more important (a_0 may be one order of magnitude over the value for FA=0 with the prolate tensor), so the conclusions arised from Section 4.5.2 are even reinforced. For example, with 51 directions and a prolate tensor with FA=0.8, even for $a_0 = 10^4$ ($A_{L,0}/\sigma_n \simeq 140$), the bias is as important as the variance; although this corresponds to the baseline image, which typically shows a high SNR, it is evident that this value may be found in many image voxels. Besides, for $A_{L,0}/\sigma_n \simeq 300$ (this SNR is an upper limit for a realistic DWI data set), the bias will be roughly 1/4 of the variance, which is clearly not negligible.

4.5.4 Discussion

It has been shown that the impact of the bias in Rician signals for WLS tensor–fitting is in general quite subtle. Notwithstanding, empirical simulations have been widely reported that reveal the clinical importance of this bias in real–world scenarios. For non–central Chi distributed signals, the bias may be a much more important source of error; the larger the number of receiving coils L, the more critical. While the variance in the estimation may be reduced increasing the number of gradient directions, this is not the case for the bias. In some cases, increasing the number of gradients will not improve the estimation, since the main source of error will be the bias and not the variance. In these cases, it may be preferable to improve the SNR by increasing NEX (unless a minimum number of gradient directions is needed for some purpose).

As an additional difficulty, the traditional WLS approach is not optimal for non–central Chi signals, since the weights commonly used are not those yielding minimum variance; although a modification to avoid this problem may be used, it makes necessary to characterize the noise power for all image voxels.

4.6 Conclusion

The importance of accounting for the impact of noise in the CMS is now evident. In particular, the bias in the **q**–space is propagated to the **R**–space producing estimation errors which cannot be removed by further regularization. In DTI, the bias has been proved a critical source of error in the case of pMRI, multiple–coils systems, or in general CMS following a non–central Chi distribution, while for Rician signals the impact is quite subtle [Bas00a, Bas06, Jon04, Sal05].

For non–Gaussian diffusion propagators, the bias of Rician–distributed DWI data sets may have a great impact even for high SNR [Cla08]. For non–central Chi statistics, it follows that the noise in the DWI might constitute an important limitation for an accurate estimation. Therefore, it will be essential to account for this factor by some means.

For DTI, WLS techniques for the estimation of the diffusion tensor may be substituted by non–linear methods such as Maximum Likelihood [Lan07], which takes into account the statistical model of noise. This strategy jointly accounts for the noise in the **q**–space and the **R**–space. It shows two main drawbacks:

- It is only valid for the Rician case, and not for non–central Chi signals; WLS remains valid for the two distributions.
- It is far more complex and slower than WLS.

In fact, one clear advantage of WLS is their general validity for any statistics of noise [Kay93]. In particular, suppose that the noise in the CMS may be modeled as additive Gaussian (this assumption is quite realistic for high SNR, see Fig. 4.1). A similar development to eq. (4.10) may be carried out to yield:

$$\log(M_L) = \log(A_L + n) = \log(A_L) + \frac{n}{A_L} - \frac{n^2}{2A_L^2} + \frac{n^3}{3A_L^3} - \cdots, \tag{4.26}$$

so the mean and variance are:

$$\begin{aligned} E\{\log(M_L)\} &= \log(A_L) - \frac{\sigma_n^2}{2A_L^2} + \mathcal{O}\left(\frac{\sigma_n^4}{A_L^4}\right); \\ \mathrm{Var}\{\log(M_L)\} &= \frac{\sigma_n^2}{A_L^2} + \mathcal{O}\left(\frac{\sigma_n^4}{A_L^4}\right), \end{aligned} \tag{4.27}$$

and the WLS problem has exactly the same formulation as in [Sal05], since the weights may be chosen exactly the same as in eq. (4.17). This result has even a deeper meaning; if an unbiased filtering stage is implemented prior to the estimation of the diffusion tensor, the residual noise in the CMS may in fact be accurately modeled as additive Gaussian, so the WLS remains valid but the propagation of the error to the **R**–space is mostly palliated: first, the SNR will be increased, and second, the residual noise will follow an approximate Gaussian distribution, so that the bias will be removed. As opposed to the approach in [Lan07], the problem is decoupled in two different problems, one in the **q**–space and the other in the **R**–space, as follows:

1. Unbiased filtering of the CMS, which is addressed in Chapter 5.
2. Estimation of the diffusion propagator assuming that the CMS is noise–free, studied in Chapters 6 and 7.

5

Noise removal in Diffusion Weighted Imaging

The particular characteristics of noise in DWI data sets were discussed in Chapter 3. The consequences of the specific nature of these signals were studied in Chapter 4, where it was shown that the noise in the images may induce large errors in the computation of the diffusion tensor or other orientation information. Moreover, these errors cannot be recovered by further regularization. At this point, the necessity of correctly accounting for the noise in DWI data sets in any estimation task should remain clear. The current Chapter deals with this topic. The first part is focused on showing the convenience of filtering DWI data sets as a pre–processing step. Although other approaches are possible, this method allows to decouple the noise removal and the estimation of fiber orientations; therefore, the denoising is independent on the estimation technique to be used, showing general validity. Amongst the available techniques in the literature, the Linear Minimum Mean Squared Error (LMMSE) filter and the Unbiased Non–Local Means (UNLM) filter are paid especial attention due to their high performance. On the other hand, it is shown that the traditional approach in DWI filtering, independently processing each gradient and baseline image, can be improved by the use of *ad hoc* techniques. DWI data sets are not only a multi–channel set of images; gathering all the channels, and the correlations between them, is the only way to account for the rich structural information provided by this modality. This is referred to as joint information, and the main focus of this Chapter is to exploit this joint information to improve the performance of denoising. Extensive experiments have been intended that evidence the goodness of this approach. The main original contribution in this Chapter is the use of joint information for DWI denoising, through the introduction of joint extensions of LMMSE and UNLM. Besides, these filters are extended to the non–central Chi case in the final part of the Chapter. In the methodological part, some insights are given about the adequacy of phantoms commonly used in the literature to assess filtering accuracy; it is shown that oversimplified phantoms may drive to erroneous conclusions in the comparison of filtering techniques, proposing a methodology to overcome this problem.

5.1 Introduction

The impact of noise in Diffusion Imaging should remain clear from Chapter 4. DWI are collected by applying a sensitizing gradient in a known direction, which produces an attenuation in the T2 (baseline) image following an exponential law (see Chapter 2 for a detailed description of the underlying physical processes). As opposed to conventional MRI, DWI data sets usually show a poor SNR: the average signal power is much lower due to diffusion attenuation, but the noise power is the same, since it is governed by the thermal disturbance in the MRI scanner (see Chapter 3). This is especially the case with modern HARDI techniques (see Section 2.7 for a brief review of them), where very strong gradients (and therefore very strong attenuation) are often applied in order to improve the contrast between different diffusion directions[1]. Moreover, the nature of noise in the CMS (Rician or non–central Chi) may render strong biases in the estimation of diffusion propagators, as shown in Chapter 4. This same biased behavior prevents the use of conventional, Gaussian-based filtering techniques.

5.1.1 State of the art on denoising of Diffusion Images

The great impact of noise in Diffusion Imaging has been widely reported previously [AF08c, Bas00a, Bas06, Jon04], and as a consequence, a number of specific techniques for noise removal in DWI/DTI have been proposed in the literature. These methods may be summarized following the classification proposed in [MF09]:

1. Regularization of the tensor fields after the diffusion tensor has been estimated [Che02, Che05, CM07, Cou04, MF04, Pen06, Pou00].
2. Estimation of the diffusion tensor (or more general orientation information) accounting for the Rician model, where regularization and tensor estimation are done simultaneously [Fil07, Pen06, Wan04b].
3. Denoising of DWI volumes previous to the estimation of the diffusion tensor or the orientation information.

It is worth noticing that the survey above is centered on DTI; for more advanced HARDI techniques, the third alternative is the most attractive, since it does not rely on any particular model. Naturally, some exceptions may be found in the literature. In [Bar08], a technique is proposed for the regularization of general displacement probability PDF fields, which may be enclosed in the first category. In [Cla08], the Rician nature of noise is accounted in some models for the estimation of these PDF, so it may be considered in the second group. However, the preferred methodology is the third one. First, the denoising process does not depend on the particular application (DTI estimation or HARDI reconstruction); second, accurate statistical models for noise, like those described in Chapter 3, are available. Additionally, the scheme presented in Section 2.5.3 holds, so that the estimation of the $\mathbf{x} \times \mathbf{q}$–space out from noisy samples and the estimation of the $\mathbf{x} \times \mathbf{R}$–space may be considered as independent problems.

This Chapter is centered on the third group of techniques, which comprise a number of approaches in the recent literature: the conventional approach [McG93], based on the properties of the second order moment of Rician data; Maximum Likelihood [Sij04] and Expectation–Maximization [Mar95] approaches; wavelets [Now99, Piž03]; unbiased Wiener filtering [MF07, MF09] or anisotropic diffusion [Ger92, Kri09] have been used. Among these approaches, the Linear Minimum Mean Squared Error (LMMSE) filter first introduced in [AF07, AF08a, AF08c] for conventional MRI, and the Non–Local Means (NLM) and Unbiased Non–Local Means (UNLM) as proposed in [AF08b, Cou08, Des08, Man08, WD07, WD08] have been shown to outperform the most common approaches in the literature.

On the other hand, it has been pointed out in [Fil07] that filtering each DWI component independently is not the best way to perform denoising, since only the combination of all the DWI can reveal the complex structure of the white matter. In other words, due to the particular nature of DWI data sets, one single DWI channel is not enough to characterize the underlying structures in the multi–dimensional

---

[1] This topic will be further discussed in Chapter 6.

image, so all the available information has to be considered as a whole. This way, not only the information of all channels, but also the correlations between them are exploited to estimate the noise–free signals. This concept has been referred to as **joint information** in [TV08].

As a final remark, all these techniques are grounded on the assumption of a Rician distribution of the noise in the CMS, since this is the most common case for single–coil MRI and most of the pMRI reconstruction algorithms. However, it was shown in Chapter 3 that two important scenarios in MRI, i.e. reconstruction from multiple array coils and the GRAPPA algorithm for pMRI, are described by non–central Chi distributions. The first part of this Chapter is focused on the description of two novel techniques for Rician noise removal in DWI, accounting for the joint information in these data sets; in the last part of the Chapter, these approaches are extended to the important non–central Chi case, thus covering a wide specter of current MRI techniques.

5.2 DWI filtering based on Rician moments

The Rician model has been the traditional keystone in MRI [Bas00a,Gud95,Jon04], as discussed in Section 3.4. The CMS can be modeled as described by eq. (3.19). However, it will be useful in this Chapter to use an alternative representation assuming that the signal of interest is real, i.e. $A = A_c + j0$. Without loss of generality, eq. (3.19) may be replaced by[2]:

$$M_i = \sqrt{(A_i + n_{c,i})^2 + n_{s,i}^2}, \tag{5.1}$$

where the subindex i has been added to denote each gradient direction. The PDF of M_i is exactly the same as in eq. (3.20). The main issue with Rician signals is that the expected value of M_i is not A_i, and yet the bias depends on A_i, so common filtering approaches based on averaging samples of M_i are not adequate. Although odd order moments of M_i have complicated expressions, the even order moments are easier to compute, and above all they represent estimates of the corresponding moments of A_i with a constant bias. For these reasons, they represent the basis for most of the filtering techniques in the literature.

In the following developments, it is useful to consider A_i itself not as a mere parameter to estimate, but also as a random variable. With this consideration, the first two even order moments of M_i can be computed from eq. (5.1). Hereafter, such moments (expected values), will be denoted as $\langle M_i^r \rangle$:

$$\begin{aligned}
\langle M_i^2 \rangle &\equiv E\{M_i^2\} = E\left\{(A_i + n_{c,i})^2 + n_{s,i}^2\right\} \\
&= E\left\{A_i^2\right\} + 2E\left\{A_i n_{c,i}\right\} + E\left\{n_{c,i}^2\right\} + E\left\{n_{s,i}^2\right\} = \langle A_i^2 \rangle + 2\sigma_n^2; \qquad (5.2) \\
\langle M_i^4 \rangle &\equiv E\{M_i^4\} = E\left\{\left((A_i + n_{c,i})^2 + n_{s,i}^2\right)^2\right\} \\
&= E\left\{A_i^4\right\} + 4E\left\{A_i^3 n_{c,i}\right\} + 2E\left\{A_i^2\left(3n_{c,i}^2 + n_{s,i}^2\right)\right\} \\
&\quad + 4E\left\{A_i\left(n_{c,i}^3 + n_{c,i} n_{s,i}^2\right)\right\} + 2E\left\{n_{c,i}^2 n_{s,i}^2\right\} + E\left\{n_{c,i}^4 + n_{s,i}^4\right\} \\
&= \langle A_i^4 \rangle + 8\langle A_i^2 \rangle \sigma_n^2 + 2\sigma_n^2\sigma_n^2 + 6\sigma_n^4 = \langle A_i^4 \rangle + 8\sigma_n^2 \langle A_i^2 \rangle + 8\sigma_n^4, \qquad (5.3)
\end{aligned}$$

since the moments of the Gaussian noises n_i ($n_{c,i}$ or $n_{s,i}$) are: $\langle n_i \rangle = 0$, $\langle n_i^2 \rangle = \sigma_n^2$, $\langle n_i^3 \rangle = 0$ and $\langle n_i^4 \rangle = 3\sigma_n^4$. In what follows two filtering techniques based on these

---

[2] The representation in terms of complex signals, $A_{c_i} + jA_{s_i}$, is completely equivalent, but the notation is more complicated. Note that this same model was used in the previous Chapter.

results are presented. They are the LMMSE filter and the UNLM filter, which as mentioned above have been successfully used in the recent literature in a number of problems.

5.2.1 The Linear Minimum Mean Squared Error filter

LMMSE estimators find estimates of $\widehat{A}_i$, which is modeled as the realization of a random variable, out from measured samples of M_i [Kay93]. This philosophy allows to include some prior information of the behavior of A_i to improve the estimation. The advantage of using a linear estimator is that it is only necessary to characterize the first and second order moments of the data to estimate. For the Rician distribution described by eq. (3.20) it is not trivial to compute the first order moment, so the strategy proposed in [AF08a, AF08c] is to estimate the squared value of A_i. The LMMSE estimator of $A_i^2(\mathbf{x})$ at each image location $\mathbf{x}$ in the domain Ω of the image reads [Kay93]:

$$\widehat{A_i^2}(\mathbf{x}) = \langle A_i^2(\mathbf{x})\rangle + C_{A_i^2M_i^2}(\mathbf{x})C^{-1}_{M_i^2M_i^2}(\mathbf{x})\left(M_i^2(\mathbf{x}) - \langle M_i^2(\mathbf{x})\rangle\right), \tag{5.4}$$

where the cross covariance $C_{A_i^2M_i^2}$ and the auto–covariance $C_{M_i^2M_i^2}$ are computed in a similar way to the moments of M_i:

$$\begin{aligned}
C_{A_i^2M_i^2}(\mathbf{x}) &= E\left\{\left(A_i^2(\mathbf{x}) - \langle A_i^2(\mathbf{x})\rangle\right)\left(M_i^2(\mathbf{x}) - \langle M_i^2(\mathbf{x})\rangle\right)\right\} \\
&= \langle M_i^4(\mathbf{x})\rangle - \langle M_i^2(\mathbf{x})\rangle^2 - 4\sigma_n^2\left(\langle M_i^2(\mathbf{x})\rangle - \sigma_n^2\right), \\
C_{M_i^2M_i^2}(\mathbf{x}) &= E\left\{\left(M_i^2(\mathbf{x}) - \langle M_i^2(\mathbf{x})\rangle\right)^2\right\} = \langle M_i^4(\mathbf{x})\rangle - \langle M_i^2(\mathbf{x})\rangle^2.
\end{aligned} \tag{5.5}$$

In practice, the moments of A_i and M_i are not known, so they have to be estimated from samples in a neighborhood $\mathcal{N}_x \in \Omega$ around the pixel at $\mathbf{x}$:

$$\langle M_i^r(\mathbf{x})\rangle \simeq \frac{1}{|\mathcal{N}_x|}\sum_{\mathbf{c}\in\mathcal{N}_x} M_i^r(\mathbf{c}) \equiv \overline{M_i^r}(\mathbf{x}), \tag{5.6}$$

where $|\mathcal{N}_x|$ is the number of samples in $\mathcal{N}_x$. For the moment $\langle A_i^2\rangle$ in eq. (5.4), the relation in eq. (5.2) may be exploited:

$$\langle A_i^2(\mathbf{x})\rangle = \langle M_i^2(\mathbf{x})\rangle - 2\sigma_n^2 \simeq \overline{M_i^2}(\mathbf{x}) - 2\sigma_n^2. \tag{5.7}$$

This is the LMMSE filter in [AF08a, AF08c]. Eq. (5.4) is applied independently for each gradient and baseline image. The estimation of the covariances in eq. (5.5) has to be repeated as well for each gradient and baseline.

5.2.2 The Unbiased Non–Local Means filter

The original NLM formulation for each gradient direction and baseline image follows [Bua05]:

$$\widehat{A}_i(\mathbf{x}) = \sum_{\mathbf{c}\in\Omega} w_i(\mathbf{x},\mathbf{c})M_i(\mathbf{c}), \tag{5.8}$$

where w_i are a set of weights computed as:

$$w_i(\mathbf{x},\mathbf{c}) = \frac{1}{U_i(\mathbf{x})}\exp\left(-\frac{d_i(\mathbf{x},\mathbf{c})}{h^2}\right),\ U_i(\mathbf{x}) = \sum_{\mathbf{c}\in\Omega}\exp\left(-\frac{d_i(\mathbf{x},\mathbf{c})}{h^2}\right), \tag{5.9}$$

where h is a parameter related to σ_n and $d_i(\mathbf{x},\mathbf{c})$ is a distance between the voxels at positions $\mathbf{x}$ and $\mathbf{c}$; instead of using a geometrical distance, NLM uses a distance in the domain of the gray levels of the image, which is:

$$d_i(\mathbf{x},\mathbf{c}) = (\mathbf{M}_i(\mathcal{N}_x) - \mathbf{M}_i(\mathcal{N}_c))^T\, G_\rho\, (\mathbf{M}_i(\mathcal{N}_x) - \mathbf{M}_i(\mathcal{N}_c)), \tag{5.10}$$

where $\mathbf{M}_i(\mathcal{N}_x)$ and $\mathbf{M}_i(\mathcal{N}_c)$ are column vectors containing the gray values of the voxels in the neighborhoods $\mathcal{N}_x$ and $\mathcal{N}_c$ of voxels $\mathbf{x}$ and $\mathbf{c}$, respectively. G_ρ is a Gaussian kernel (diagonal matrix) which gives a higher weight to the voxels of the neighborhood geometrically closer to the center of $\mathcal{N}_x$. The computational load associated to eq. (5.8) is prohibitive, so the domain Ω is usually substituted by a neighborhood $\mathcal{N}_x'$ of voxel $\mathbf{x}$. Besides, it is proposed in [Man08] to change the weight $w(\mathbf{x},\mathbf{x})$ in eq. (5.9) by the maximum of $w(\mathbf{x},\mathbf{c})$, $\mathbf{c}\neq\mathbf{x}$ to avoid over–weighting the central voxel of $\mathcal{N}_x'$. A similar procedure is applied to the center of G_ρ.

The optimality of NLM for additive and multiplicative noise has been shown in [Bua05]. For Rician noise, the average of noisy samples $M_i(\mathbf{c})$ would introduce a bias in the estimation [Xu08]. Like for LMMSE, the squared value of A_i may be estimated instead[3] [Des08, WD08], so eq. (5.8) reads:

$$\widehat{A_i^2}(\mathbf{x}) = \sum_{\mathbf{c}\in\Omega} w_i(\mathbf{x},\mathbf{c}) M_i^2(\mathbf{c}) - 2\sigma_n^2. \tag{5.11}$$

The second order moment $\langle M_i^2(\mathbf{x})\rangle$ is estimated as a non–local sample average, and eq. (5.2) is used to compute the estimation $\widehat{A_i^2}$ as the non–local second order moment $\langle A_i^2(\mathbf{x})\rangle$. This is the so–called UNLM filter, for which each gradient image and baseline are filtered independently (like for LMMSE).

5.3 Exploiting joint information: filters in the $\mathbf{x}\times\mathbf{q}$–space

The filtering techniques reviewed in Sections 5.2.1 and 5.2.2 process each gradient image (and baseline) as an independent image. This methodology does not allow to exploit the correlations (joint information) present in diffusion data sets: corresponding voxels of different volumes correspond to the same anatomical structure, so gathering the information of all the gradient images as a whole might be useful to better model image properties and improve the estimation. This way, the filtering is not performed in the 3D $\mathbf{x}$–space, but instead the dimensionality is extended to the $\mathbf{x}\times\mathbf{q}$–space. Hence, neighborhoods are not limited to the space domain, comprising proximal points in the $\mathbf{q}$–space, together with their statistical correlations. In the remainder of this Section, the LMMSE filter and the UNLM filter are extended to meet this requirement.

5.3.1 Notation

The vectors $\mathbf{M}^2$ and $\mathbf{A}^2$ denote the set of noisy measurements $[M_0^2,\ldots,M_{Z-1}^2]^T$ and the set of noise–free values to estimate $[A_0^2,\ldots,A_{Z-1}^2]^T$, respectively. The spatial dependency $\mathbf{M}^2(\mathbf{x})$ and $\mathbf{A}^2(\mathbf{x})$ will be made explicit were needed and dropped down elsewhere. Note that the set of indices $i=0,\ldots,Z-1$ comprise both gradient images and baselines: $M_0^2,\ldots,M_{Z_b-1}^2$ are the Z_b (squared) baselines, and $M_{Z_b}^2,\ldots,M_{Z-1}^2$ are the

---

[3] Other approaches based on the first order moment of the Rician distribution are possible [AF08b]. In this work, an analytical correction of the bias, dependent on the estimated value of A_i, has to be computed. Only the traditional methodology based on even order moments [McG93] is discussed here.

Z_g (squared) gradients. Therefore, $Z = Z_b + Z_g$, where $Z_b < Z_g$. For convenience, the following auxiliary set of indices for each integer N, $0 < N \leq Z_g$, are defined:

$$\Theta_i^N = \begin{cases} \{j \mid 0 \leq j < Z_b\}, & \text{if } 0 \leq i < Z_b \text{ (baselines).} \\ \text{The set of indices of the } N \text{ clos-} \\ \text{est gradient directions to } g_i. \end{array}, & \text{if } Z_b \leq i < Z \text{ (gradients).} \end{cases} \tag{5.12}$$

If the index i corresponds to a baseline, Θ_i^N is the whole set of baseline indices $\{0,\ldots,Z_b-1\}$ for any value of N. If i corresponds to a gradient image, Θ_i^N is the set of the N indices corresponding to the gradient directions most similar to the i–th direction. For $N=1$, Θ_i^N reduces to i. For $N = Z_g$, $\Theta_i^N = \{Z_b,\ldots,Z-1\}$. Accordingly, $\mathbf{A}_{i|N}$ denotes the vector of components of $\mathbf{A}$ corresponding to the indices in Θ_i^N.

The expected value of $\mathbf{A}(\mathbf{x})$ is denoted $\langle\mathbf{A}(\mathbf{x})\rangle$, and its sample moment in a neighborhood $\mathscr{N}_x \subset \Omega$ of $\mathbf{x}$ reads $\overline{\mathbf{A}}(\mathbf{x})$. $\Omega \subset \mathbb{R}^3$ is the domain of the image (the $\mathbf{x}$–space). Finally, the cross covariance matrix $\mathbf{C}_{A^2M^2}(\mathbf{x})$, and the auto covariance matrix $\mathbf{C}_{M^2M^2}(\mathbf{x})$ are defined as:

$$\begin{aligned} \mathbf{C}_{A^2M^2}(\mathbf{x}) &= E\left\{\left(\mathbf{A}^2(\mathbf{x}) - \langle\mathbf{A}^2(\mathbf{x})\rangle\right)\left(\mathbf{M}^2(\mathbf{x}) - \langle\mathbf{M}^2(\mathbf{x})\rangle\right)^T\right\}. \\ \mathbf{C}_{M^2M^2}(\mathbf{x}) &= E\left\{\left(\mathbf{M}^2(\mathbf{x}) - \langle\mathbf{M}^2(\mathbf{x})\rangle\right)\left(\mathbf{M}^2(\mathbf{x}) - \langle\mathbf{M}^2(\mathbf{x})\rangle\right)^T\right\}. \end{aligned} \tag{5.13}$$

5.3.2 Multi–channel LMMSE based on joint information

Generalizing eq. (5.4) to the vector case is straightforward [Kay93]:

$$\widehat{\mathbf{A}^2}(\mathbf{x}) = \langle\mathbf{A}^2(\mathbf{x})\rangle + \mathbf{C}_{A^2M^2}(\mathbf{x})\mathbf{C}_{M^2M^2}(\mathbf{x})^{-1}\left(\mathbf{M}^2(\mathbf{x}) - \langle\mathbf{M}^2(\mathbf{x})\rangle\right). \tag{5.14}$$

In this approach, one single estimator is defined for all the gradient and baseline images as a whole. The moments $\langle\mathbf{M}^2\rangle$ and $\langle\mathbf{A}^2\rangle$ are estimated from samples directly extrapolating eqs. (5.6) and (5.7) for each component of $\mathbf{M}^2$ and $\mathbf{A}^2$. The estimation of $\mathbf{C}_{A^2M^2}$ and $\mathbf{C}_{M^2M^2}$ is more involved, since it requires the computation of all crossed moments $E\{(X_i - \langle X_i\rangle)(Y_j - \langle Y_j\rangle)\}$, see eq. (5.13). This computation heavily increases the computational load, which is $O(Z^2)$. An alternative approach has been proposed in [TV08], which is included in Appendix B. Under some assumptions, the following equivalences hold:

$$\begin{aligned} \mathbf{C}_{A^2M^2} &= \varsigma\langle\mathbf{A}^2\rangle\langle\mathbf{A}^2\rangle^T; \\ \mathbf{C}_{M^2M^2} &= \varsigma\langle\mathbf{A}^2\rangle\langle\mathbf{A}^2\rangle^T + 4\sigma_n^2\mathrm{diag}\left(\langle\mathbf{A}^2\rangle\right) + 4\sigma_n^4\mathrm{I}_Z, \end{aligned} \tag{5.15}$$

where:

$$\varsigma = \frac{\langle A_\beta^4\rangle - \langle A_\beta^2\rangle^2}{\langle A_\beta^2\rangle^2} > 0, \tag{5.16}$$

is the normalized local structural variability measured for any of the baseline images β, so $0 \leq \beta < Z_b$. I_Z is the identity matrix, and $\langle A_\beta^4\rangle$ may be estimated from $\overline{M_\beta^4}$ using eq. (5.7). Note that the computation of the matrices in eq. (5.15) requires only to estimate the second–order moment of each component of $\mathbf{A}^2$, so the complexity is only $O(Z)$. On the other hand, the inversion of $\mathbf{C}_{M^2M^2}$ at each image location $\mathbf{x}$ is a clear computational burden. Instead, it has been proved in [TV08] that the following recursive approximation may be used in practically all image voxels:

$$\begin{aligned} \mathbf{C}_{M^2M^2}^{-1}\left(\mathbf{M}^2 - \langle\mathbf{M}^2\rangle\right) &\simeq \mathbf{w}^{\mathscr{O}}; \\ \mathbf{w}^{l+1} &= \left(\mathbf{M}^2 - \langle\mathbf{M}^2\rangle\right) - 4\sigma_n^2\widetilde{\mathbf{C}}_{M^2M^2}^{-1}\mathbf{w}^l; \\ \mathbf{w}^0 &= \left(\mathbf{M}^2 - \langle\mathbf{M}^2\rangle\right); \\ \widetilde{\mathbf{C}}_{M^2M^2}^{-1} &= \eta\mathbf{1}\mathbf{1}^T + \mathrm{diag}(\mathbf{e}), \end{aligned} \tag{5.17}$$

where η is a scalar constant and $\mathbf{1},\mathbf{e}$ are $Z\times 1$ vectors defined by:

$$\eta = -\left(4\sigma_n^2\left(\frac{4\sigma_n^2}{\varsigma}+\sum_{i=0}^{L}\langle A_i^2\rangle\right)\right)^{-1};\quad \mathbf{e}_i = \left(4\sigma_n^2\langle A_i^2\rangle\right)^{-1};\quad \mathbf{1}_i = 1,$$

see Appendix B for details. In practice, it is enough to use only one term in the recursion, so $\mathscr{O}=1$. Besides, the products with $\widetilde{\mathbf{C}}^{-1}_{M^2M^2}$ are very efficient to compute due to its simple structure (it is easy to check that the complexity remains $O(Z)$).

Eqs. (5.15) and (5.17), together with the estimation procedure described for the moments of $\mathbf{M}^2$ and $\mathbf{A}^2$, may be casted into eq. (5.14) to perform the joint LMMSE filtering. The main advantage of this approach compared to the LMMSE described in Section 5.2.1 is that it accounts for the joint information of the whole data set: to estimate the value of $A_i(\mathbf{x})$, not only the measurements $M_i(\mathbf{c}),\mathbf{c}\in\mathscr{N}_x$ are considered, but all the measurements $M_j(\mathbf{c}),\mathbf{c}\in\mathscr{N}_x$ of other channels as well. Besides, the covariance matrices $\mathbf{C}_{M^2M^2}$ and $\mathbf{C}_{A^2M^2}$ account for the correlations between different channels. Although the joint LMMSE has been shown to outperform LMMSE, some ringing artifacts and over–blurring may appear under some circumstances [TV08]. To further study this behavior, eq. (5.14) may be re–written explicitly using eqs. (5.15) and (5.17):

$$\widehat{\mathbf{A}^2} = \langle\mathbf{A}^2\rangle + \varsigma\langle\mathbf{A}^2\rangle\langle\mathbf{A}^2\rangle^T\,\mathbf{Q} = \langle\mathbf{A}^2\rangle + \langle\mathbf{A}^2\rangle\cdot\chi, \tag{5.18}$$

for:

$$\begin{aligned}\mathbf{Q} &= \mathbf{C}^{-1}_{M^2M^2}\left(\mathbf{M}^2-\langle\mathbf{M}^2\rangle\right);\\ \chi &= \varsigma\langle\mathbf{A}^2\rangle^T\,\mathbf{Q},\end{aligned}$$

where $\mathbf{Q}$ is the column vector recursively computed in eq. (5.17), and χ is its inner product with $\varsigma\langle\mathbf{A}^2\rangle$. The moments in eq. (5.18) are computed as local sample averages: $\langle\mathbf{A}^2\rangle\simeq\overline{\mathbf{M}^2}-2\sigma_n^2$. The estimated value $\widehat{\mathbf{A}^2}$ is thus computed as the low–pass filtered $\mathbf{M}^2-2\sigma_n^2$, and further corrected with the term $\langle\mathbf{A}^2\rangle\cdot\chi$, which is proportional to ς:

- If the local variability of the image is small, ς vanishes and the estimated value is approximately the (unbiased) local average of the image, $\overline{\mathbf{M}^2}-2\sigma_n^2$.
- If the local variability is higher, this local average is corrected with the term $\langle\mathbf{A}^2\rangle\cdot\chi$ to avoid the low–pass filtering and preserve the contours.

The difference with the conventional LMMSE is that the factor χ is the same for all baselines and gradient images, and it is fixed taking into account the joint information of all channels. This technique has two main drawbacks, which are responsible of the aforementioned artifacts:

1. The correction factor χ is the same for the baselines as well as for all the gradient images, which show very different properties depending on the tissue.
2. In the presence of strong eddy current distortion, a corresponding voxel in two different gradient images does not necessarily correspond to the same physical point [And02,Net04,Nie04], introducing an important source of error in the joint model.

To overcome these problems, it can be considered that the baseline images, if more than one, have identical properties. For the gradients, images corresponding to similar directions have similar properties, and are affected in a similar way by eddy currents. Thus, the filter LMMSE-N is defined as:

$$\begin{aligned}\widehat{A_i^2} &= \langle A_i^2\rangle + \varsigma\langle A_i^2\rangle\left(\langle \mathbf{A}^2{}_{i|N}\rangle^T \mathbf{C}^{-1}_{M^2_{i|N}M^2_{i|N}}\left(\mathbf{M}^2_{i|N} - \langle \mathbf{M}^2{}_{i|N}\rangle\right)\right)\\ &= \langle A_i^2\rangle + \chi_{i|N}\langle A_i^2\rangle, \qquad (5.19)\end{aligned}$$

Eq. (5.19) is formally very similar to eq. (5.14), but now the estimation is done in a different way for each channel. Only the joint information shared by the channels in Θ_i^N and their correlations are accounted to filter the channel i. This way, only volumes with similar characteristics are filtered together. For the **baselines**, the correction $\chi_{i|N}$ is computed only once, since Θ_i^N is the same for all of them (the whole set $\{0,\ldots,Z_b-1\}$). For the **gradient images**:

- If $N < Z_g$ the correction $\chi_{i|N}$ is computed for each gradient, because each set Θ_i^N is different from each other. Since there is a set of size N for each of the Z_g gradients, the complexity is $O(NZ_g)$.
- If $N = Z_g$, all Θ_i^N are the same (the whole set $\{Z_b,\ldots,Z-1\}$), so $\chi_{i|N}$ is computed once and the complexity is only $O(Z_g)$.

Under a different point of view, a neighborhood in the $\mathbf{q}$–space is defined, so that only the points which are proximal in this space are considered to process each voxel. Given this idea of proximity, both in the space domain and in the domain of the gradients, LMMSE–N may be considered to work in the $\mathbf{x} \times \mathbf{q}$–space.

5.3.3 Multi–channel UNLM based on joint information

In order to adapt the same idea behind LMMSE–N, the notion of neighborhood in the $\mathbf{q}$–space is to be used; assuming once again that similar gradient directions show similar behaviors, the UNLM–N filter is defined as:

$$\widehat{A_i^2}(\mathbf{x}) = \sum_{j\in\Theta_i^N}\sum_{\mathbf{c}\in\mathcal{N}_x'} w_i^j(\mathbf{x},\mathbf{c})M_j^2(\mathbf{c}) - 2\sigma_n^2, \qquad (5.20)$$

where now:

$$w_i^j(\mathbf{x},\mathbf{c}) = \frac{1}{U_i(\mathbf{x})}\exp\left(-\frac{d_i^j(\mathbf{x},\mathbf{c})}{h^2}\right),\; U_i(\mathbf{x}) = \sum_{j\in\Theta_i^N}\sum_{\mathbf{c}\in\mathcal{N}_x'}\exp\left(-\frac{d_i^j(\mathbf{x},\mathbf{c})}{h^2}\right), \qquad (5.21)$$

and:

$$d_i^j(\mathbf{x},\mathbf{c}) = \left(\mathbf{M}_i(\mathcal{N}_x) - \mathbf{M}_j(\mathcal{N}_c)\right)^T G_\rho\left(\mathbf{M}_i(\mathcal{N}_x) - \mathbf{M}_j(\mathcal{N}_c)\right). \qquad (5.22)$$

UNLM works by averaging voxels with weights w found by correlating patches around the voxels of interest. UNLM–N looks for these correlated patches not only in the same gradient image i, but also in gradient images j near the direction i. This way, correlations between the DWI channels (i.e. the joint information) are exploited. Although this is not the first attempt to adapt UNLM to a multi–channel environment [WD07], the former work shows a very poor performance. In the approach here presented the non–local behavior is reduced to the spatial domain, but in the $\mathbf{q}$–space local neighborhoods are considered instead to avoid artifacts. Like LMMSE–N, UNLM–N works by averaging voxels in neighborhoods in the $\mathbf{x} \times \mathbf{q}$–space; this neighborhoods are non–local in the the spatial domain (where distances in the gray levels are considered) and local in the domain of the gradients (where geometrical distances are used).

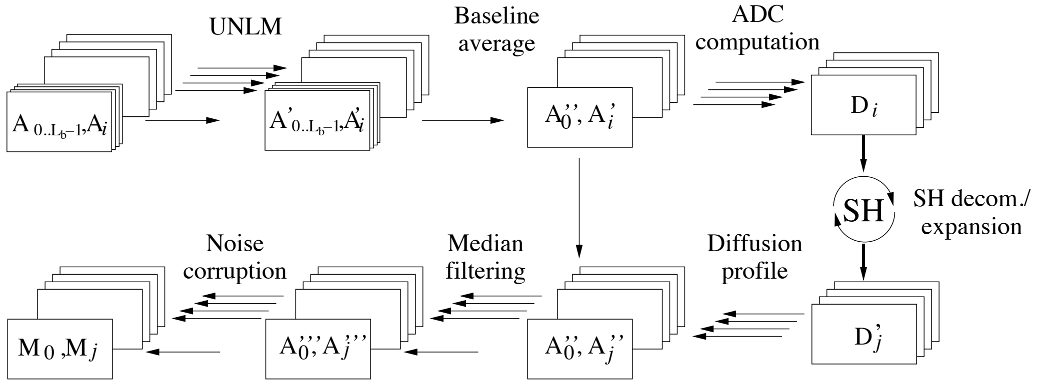


Figure 5.1: Overview of the construction of the synthetic phantom. The input is the original (real) DWI volume, and the output the synthetic volume.

5.4 Performance of the filtering algorithms

5.4.1 Design of phantom data for quantitative evaluation

The quantitative assessment of filtering performance has been carried out in different ways. They include visual assessment [AF08c, TV08, WD07, WD08] and indirect measures based on the properties of DTI volumes recovered from filtered DWI [AF08c, MF09]; in this case, the computation of related parameters, such as the smoothness in the fiber tracts estimated from DTI [MF09], makes it difficult to evaluate the filtering performance, since the final result depends on a number of factors others than the filtering itself. On the other hand, direct evaluation on filtered DWI is a difficult task; in [WD07], 12 acquisitions of the same patient are available, so a leave–one–out strategy is used to filter one of the volumes each time and compare the result to a *noise–free* image obtained from the remaining 11 volumes. When multiple observations are not available, phantom data has to be used. As opposed to conventional MRI, for which very realistic phantoms exist and have been intensively used [Col98], there is not a DWI synthetic data set widely accepted as a standard, so over-simplified models are used like in [AF08c, TV08, WD08], based on large homogeneous regions simulating coarse crossing fiber tracts. Commonly used configurations are crossing sections in two [AF08c, MF09, WD08] and three dimensions [TV08], although other approaches like the *earth* and *logarithm* in [MF09], or the logarithmic spiral in [AF08c] are possible. These simple configurations are not representative of the complex architecture of the white matter of the brain, which in general presents multiple fiber crossings, bending and sharp changes in curvature/orientation in the space of a few voxels. On the other hand, a classical methodology to obtain phantom data with image characteristics similar to real data may be:

1. Pick up a real data set and denoise/regularize it by some means. This will be the golden standard.
2. Corrupt the regularized data set with a **known** random process.
3. Filter the noisy data set, and compare it with the ground truth.

In order to adapt this methodology to the DWI case, the process depicted in Fig. 5.1 is followed [TV09a]. In point (1), a real DWI data set with 8 baselines $A_0,\ldots,A_7$ and 51 gradient directions $A_8,\ldots,A_{58}$ is considered. To that end, volume BWH3 has been selected, since it has a good spatial resolution, enough gradient directions to resolve complex micro–architectures, and yet a small enough b–value to keep a high SNR. The denoising/regularization proceeds in three stages:

Stage 1 UNLM denoising for each DWI channel and baseline average.

Stage 2 Regularization of the DWI data set across the gradient directions.

Stage 3 Median filtering.

These techniques are further explained next.

UNLM denoising of each channel

UNLM–1 (i.e., the conventional UNLM) is chosen as a well tested algorithm, and as well to avoid any bias which could favor either LMMSE–N or UNLM–N. Assuming that the residual noise in the images is close to Gaussian, or at least symmetrically distributed, the eight filtered baselines are averaged to achieve a nearly *noise–free* baseline A_0''. The gradient images are renumbered $A_1', \dots A_{51}'$.

Regularization of the DWI data set across the gradient directions

According to [Fil07], a correct regularization of a DWI data set must account for the relations between the images in different gradient directions. While stage 1 is centered on the spatial regularization of all channels (i.e. in the **x**–space), in stage 2 the regularization is performed in the **q**–space. To that end, the technique proposed in [Des06, Des07] is used. Going back to eq. (2.39), the Apparent Diffusion Coefficient (ADC) may be expressed with total generality in the form:

$$D_i \equiv D(b, \mathbf{g}_i) = -\frac{1}{b}\log\left(\frac{A_i'}{A_0''}\right) \quad i = 1\dots51, \tag{5.23}$$

where $\mathbf{g}_i$ is the i–th unitary gradient direction, which may be expressed in terms of its spherical coordinates –in physics convention– (θ_i, ϕ_i). D_i are a sampled noisy version of the ADC at the measured b–value. To regularize this function (which is defined in the sphere of radius $q = \sqrt{b/4\pi\tau}$ belonging to the **q**–space), it has been proposed in [Des06] to expand it in the basis of Spherical Harmonics (SH)[4]:

$$D'(b, \mathbf{g}) = \sum_{l=0}^{3} \sum_{m=-2l}^{2l} C_{2l}^{m} Y_{2l}^{m}(\mathbf{g}), \tag{5.24}$$

where Y_l^m are the basis functions and C_l^m are the coefficients of the expansion, fitted to data by LS. The regularization is achieved by two means:

1. Truncating the SH expansion to order $2l = 6$.
2. Introducing a Tikhonov regularization term in the LS fitting, which is the parameter λ used in [Des06] (its suggested value is $\lambda = 0.006$).

---

[4] SH have been of paramount importance in the development of diffusion imaging based on HARDI. The reader is referred to Section 6.4 and Appendix D for a detailed mathematical description of this tool. At this point, it is enough to say that SH form an orthonormal basis for all those functions defined on the sphere of radius 1 (which is to say, all those functions taking values for the gradient directions $\mathbf{g}_i$). Basis functions of lower orders l correspond to lower energies of the second order angular derivatives, and therefore to slow variations in the ADC profile.

Eq. (5.24) represents the ADC for any direction $\mathbf{g}$, so the regularized gradients may be computed for an arbitrary gradient direction using eq. (2.39):

$$A''_j \equiv A''\left(\mathbf{g}'_j\right) = A''_0 \exp\left(-b'D'\left(b,\mathbf{g}'_j\right)\right), \tag{5.25}$$

where a more usual b–value for practical applications has been chosen: $b' = 1200$ s/mm^{-2}. This election serves to increase the contrast between each DWI volume[5]. Two data sets have been generated using eq. (5.25):

- *ph15*, with 15 gradient directions $\mathbf{g}'_j$ taken from CDR1.
- *ph51*, with 51 gradient directions $\mathbf{g}'_j$ taken from BWH2.

Median filtering

Due to a poor fit of SH, some outliers may appear in the background of the image or in the ventricles. Although these zones will be avoided in the computation of the similarity measures (see Section 5.4.2), median filtering is carried out to achieve a visually adequate result. The data set is median filtered slice by slice and gradient by gradient with a mask of size 3×3.

Noise corruption

The point 2 in the methodology described is to corrupt the phantom with Rician noise:

$$M_j = \sqrt{\left(A'''_j + n_c\right)^2 + n_s^2},\ j = 0\dots Z'_g, \tag{5.26}$$

where n_c and n_s are independent Gaussian random processes with zero mean and variance σ_n^2. Spatially correlated noises $n_{c,s}$, instead of white noise, are generated to simulate realistic scanning conditions. From white Gaussian noises $\tilde{n}_{c,s}$ with noise power σ_n^2, the following transformation is performed:

$$n_{c,s}(k_1,k_2) = \left(\sum_{p_1,p_2} G^2(p_1,p_2)\right)^{-1} \sum_{p_1,p_2} \tilde{n}_{c,s}(p_1,p_2)G(k_1-p_1,k_2-p_2), \tag{5.27}$$

where G is a Gaussian kernel with isotropic variance η^2. The signal is divided by the energy of G to keep its power constant.

The results for an arbitrary gradient direction of the noise–free phantom (*ph15*), the noise–corrupted phantom for PSNR=41.9 ($\sigma_n = 200$) and PSNR=53.8 ($\sigma_n = 50$), together with an original slice of V_3 may be seen in Fig. 5.2. As a final comment, it is important to stress that the data set proposed **is not** intended as a realistic anatomical phantom, but only as a noise–free phantom to assess the performance of filtering. The edge configurations and image structures are similar to what may be found in a real data set, but obviously not the underlying neural architectures or connections.

5.4.2 Filtering performance criteria based on ground truth

The point 3 in the methodology used is to compare the filtered image with the noise–free phantom. To that end, two different criteria are used:

---

[5] Note that the ADC is assumed to have a similar shape both for b and b': $D(b,\mathbf{g}) \simeq D(b',\mathbf{g})$. This assumption, which is the keystone for a number of HARDI techniques [Des06, Öza06], will be further discussed in Chapters 6 and 7.

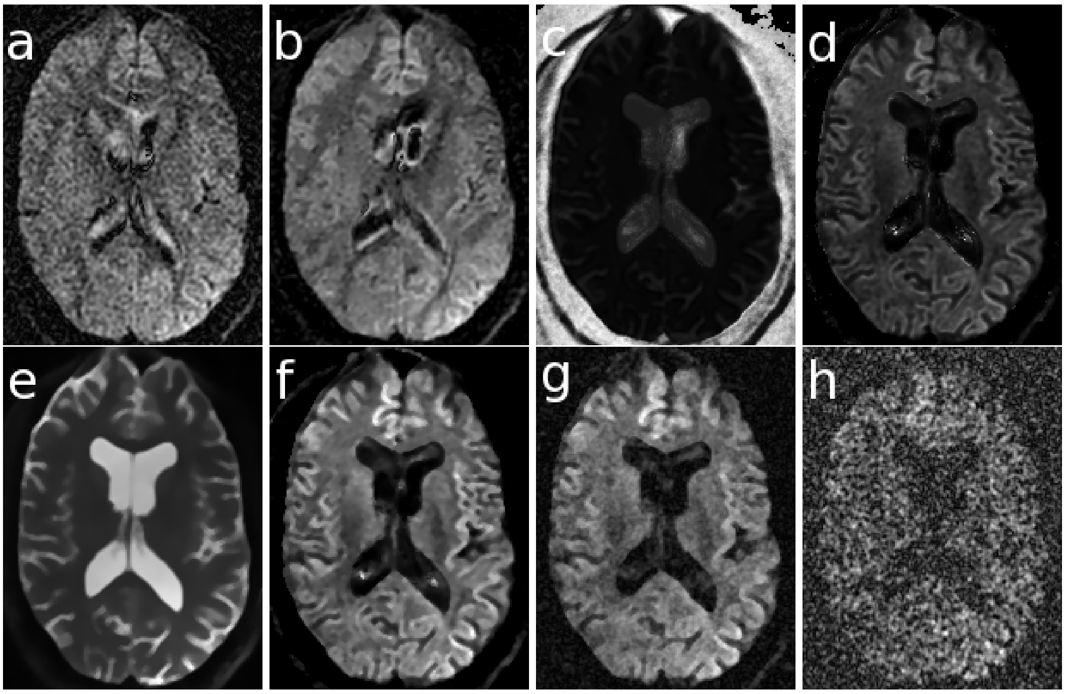


Figure 5.2: Central slices of the DWI phantom for an arbitrary gradient direction. From the original gradients A_i (**a**), UNLM is used to regularize the volumes in the $\mathbf{x}$–space (A'_i in **b**). Then the ADC (D_i) is computed and regularized in the $\mathbf{q}$–space (D'_j) via SH decomposition, with a total of 28 coefficients C^m_{2l} (**c** shows the first one for $l = 0$). Arbitrary gradient directions A''_j are generated with a higher b–value (**d**). Finally, a median filter is used to obtain the *noise–free* phantom A'''_j, comprising the averaged baseline A'''_0 (**e**) and each of the 15 gradient images A'''_j (**f**). Correlated noise is added to produce the noisy phantoms M_i, with a maximum PSNR (see Section 5.4.2) of 53.8dB (SNR=17.6dB, **g**) and a minimum of 41.9dB (SNR=5.6dB, **h**).

- The Peak Signal to Noise Ratio (**PSNR**) has been commonly used in the recent literature as an intuitive an easy to compute measure [Cou08, WD08]. It is defined as:

  $$\mathrm{PSNR} = 20\log_{10}\frac{\mathscr{A}}{\mathrm{RMSE}}, \tag{5.28}$$

  with RMSE the Root Mean Squared Error and $\mathscr{A} = \max_{\mathbf{x}} A(\mathbf{x})$ the peak amplitude of the signal, which is 25,000 for the phantom data presented in Section 5.4.1. Other RMSE–related measures have been used in the literature [AF08c, TV08, WD07], such as the SNR. The SNR is proportional to the PSNR for a given signal power, so they may be considered equivalent.

- The problem with the PSNR is that it represents too rough a measure of similarity, since it does not account for the similarity between image structures, but only for the similarity between gray levels. To overcome this problem, the Structural Similarity (**SSIM**) index introduced in [Wan04a] is used.

Although some other measures related to tensor data, such as the similarity in the Fractional Anisotropy (FA) or the Mean Diffusivity (MD) might be thought of, their use has two inherent drawbacks:

1. The estimation of the tensor introduces an extra regularization, since only 6 degrees of freedom are estimated from a typically larger number of gradient directions. This way, it is difficult to distinguish if a given result is the effect of the filtering process, of the estimation process, or a combination of both. Since the interest is only in the evaluation of filtering performance, it is preferable to isolate the former effect.

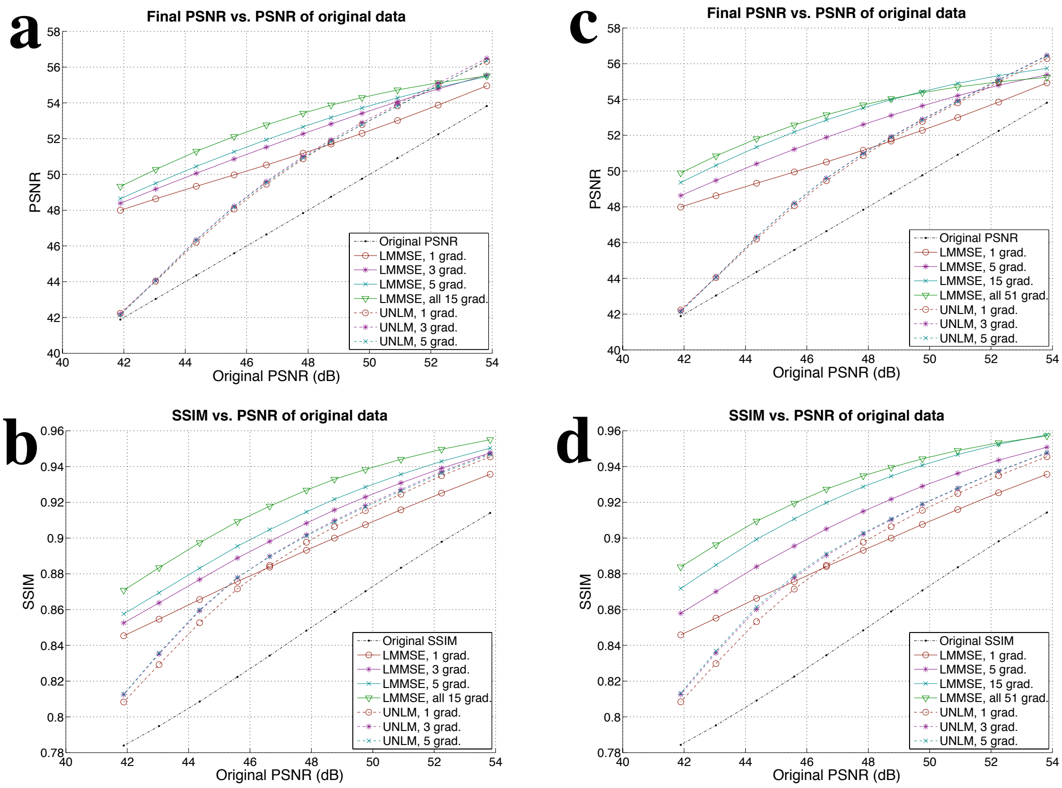


Figure 5.3: Performance indices as a function of the PSNR of the original data, for *ph15* (a and b), and *ph51* (c and d): PSNR (a and b), and SSIM (b and d).

2. As stated in Section 5.4.1, the phantom data considered is only intended to simulate the characteristics of the images but not the anatomical properties of real data. Evaluating tensor (anatomical) measures would not be conclusive, since the anatomical properties of the phantom are not realistic.

As a final comment, the indices are computed only in the structures of interest inside the image, so the background is removed by automatic thresholding [Ots79]. All the measurements reported refer to mean values amongst the DWI without the baseline images. The reason is that the SNR in the baseline is much greater, so including it in the computation could reduce the significance of the results.

5.4.3 Quantitative performance assessment based on ground truth

The LMMSE and UNLM approaches are quantitatively compared and the improvement of using joint information of DWI is measured. As mentioned above, the direct evaluation of filtering performance over DWI data sets before DTI or HARDI reconstruction allows to isolate the behavior (and possible artifacts) due to the filtering from the behavior (and possible artifacts) due to the regularization in the estimation. Identical methodologies have been commonly used for this task in the literature [AF08c, Bas06, Cou08, MF09, WD07, WD08]. The phantom data presented in Section 5.4.1 is used together with the performance criteria in Section 5.4.2 to compare several configurations of LMMSE–N and UNLM–N. For LMMSE, since the inter–slice distance of the phantom is not very large, the window size is set to $5 \times 5 \times 3$. Four different configurations are tested depending on the scenario:

- For *ph15*, LMMSE–1, LMMSE–3, LMMSE–5 and LMMSE–15.

$\mathcal{N}_x$	$\mathcal{N}_x'$	ρ	h
$3\times 3\times 3$	$7\times 7\times 3$	1	$h=\sigma_n$

Table 5.1: Summary of the parameters used for UNLM–N in all experiments.

- For *ph51*, LMMSE–1, LMMSE–5, LMMSE–15 and LMMSE–51.

For UNLM, the parameters used in all the experiments reported are summarized in Table 5.1. The values suggested in [Man08] are mostly used (the sizes of the neighborhoods are adapted to a 3D image), except for h, which is set to a typical value for DWI [Cou08, WD07]. Only UNLM–1, UNLM–3 and UNLM–5 are tested in both scenarios, since using more gradients becomes computationally prohibitive. A range of PSNR between 41.9 dB and 53.8 dB has been used, which at the sight of Fig. 5.2 seems reasonable. The results shown in Fig. 5.3 allow to extract the following conclusions:

1. The two performance indices yield consistent results over a wide range of input PSNR.
2. The joint model performs clearly better for LMMSE; the more joint information, the higher indices. This behavior holds both for 15 and 51 gradient directions.
3. The use of joint information is more profitable for more gradient directions. This is especially noticeable for the SSIM index (compare b and d). Exploiting the correlations between different channels allows to better resolve the structures of the image, which is especially useful to preserve the contours.
4. The improvement with the joint model for UNLM is only marginal, but at least the results are not worsened (as opposed to the method in [WD07]).
5. UNLM–N performs better than the conventional LMMSE–1 for medium and high PSNR. But LMMSE–N, even for $N=3$ or 5, outperforms all configurations of UNLM–N for medium and low SNR.

Summarizing this Section, the advantage of using joint information for DWI filtering remains clear. The more joint information, the greater the improvement. This is especially the case for LMMSE–N, although UNLM–N may benefit from this technique as well. Comparing these two filters, LMMSE–N seems preferable for medium–low PSNR, but its advantage over UNLM–N is not so clear for medium–high PSNR. LMMSE–N is thus preferable in scenarios of low SNR with a large number of gradient directions, which makes it advisable for HARDI scenarios (this assertion will be further validated later on).

5.4.4 On the use of phantom data

The aim of this Section is to stress the importance of using an adequate phantom for performance assessment, as opposed to the conventional crossing sections phantoms used in the literature. To that end, analogous experiments to those in Section 5.4.3 are presented here, but now the 3D phantom described in [TV08], and sketched in Fig. 5.4, is used. It consists of a sphere with radius $R=120$ inside a black background. The baseline image has a decreasing magnitude towards the borders: $A_0=230/(1+(r/200)^2)$, except for a band with constant magnitude $A_0=255$ for $|x|<35$. There are three fiber bundles along each coordinate axis, with diffusion directions following them: for $|y,z|<35$, the eigenvalues are $\lambda=$

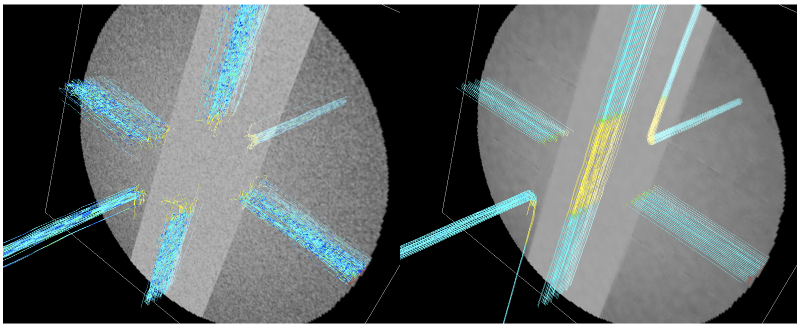


Figure 5.4: Synthetic phantom based on crossing sections. The baseline image, together with some representative fiber bundles, is shown to give an idea of its configuration. Left: noisy phantom. Right: LMMSE–15 filtered phantom.

$[1, 0.2, 0.2] \cdot 10^{-3}$ mm^2/s, and so on. The rest of the phantom is isotropic with $\lambda^j = 0.25 \cdot 10^{-3}$ mm^2/s, and b=1200 s/mm^2; 15 gradient directions have been used. For the sake of comparison, UNLM–1 and LMMSE–15 are tested. Additionally, the Conventional Approach (CA) in [McG93] is included in the experiment. It is implemented as a Gaussian low–pass filter with isotropic standard deviation of $\eta = 1.5$ voxels, which is applied to the squared signal M_i^2 and the bias removed using eq. (5.2). PSNR between 48dB and 54dB (medium and high SNR range) are tested in Fig. 5.5, comparing the results for the synthetic phantom of Fig. 5.4 and for the phantom data introduced in Section 5.4.1. The following comments arise from this experiment:

1. The two indices yield contradictory results with the cross sections phantom, as opposed to the phantom data used in Section 5.4.3.
2. For PSNR, CA seems to outperform UNLM for medium SNR with the cross sections phantom. CA is not intended to preserve image contours (it is only a low–pass filter), so this behavior does not seem reasonable. This is not the case with the phantom proposed.
3. For SSIM, CA seems to outperform LMMSE for practically all SNR with the cross sections phantom. The same comment in the previous point holds now.

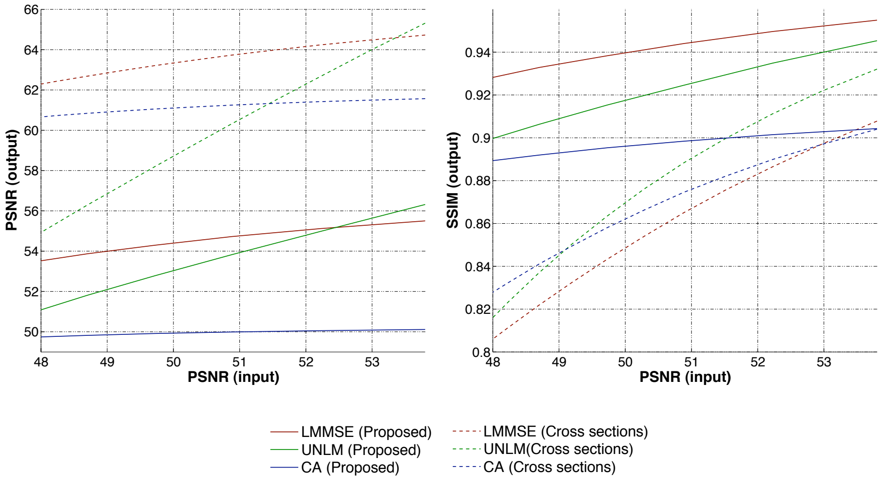


Figure 5.5: Filtering performance with the phantom data used (solid) and with a conventional phantom (dashes), for LMMSE–15 (red), UNLM (green) and the Conventional Approach (blue). The PSNR (left) and the SSIM (right) are shown for 15 gradient directions.

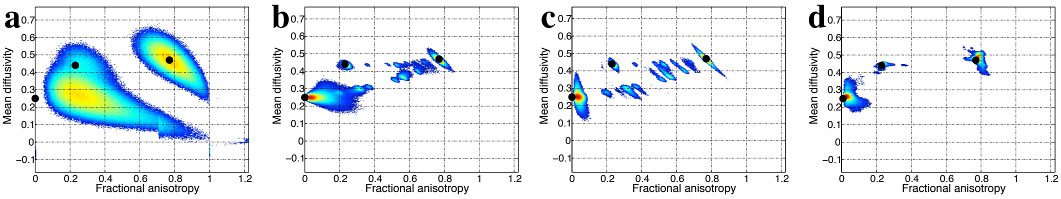


Figure 5.6: Log–plots of the 2-D histograms in the FA/MD plane (MD is scaled by 1000 for convenience) for the phantom of Fig. 5.4: (a) original noisy data, (b) LMMSE–5, (c) LMMSE–15 and (d) UNLM–1. The ideal centroids for each region are represented as black spots. Those pairs with a frequency minor than 10^{-3} times the maximum one have been removed from the plot.

4. Comparing UNLM and LMMSE, the results for PSNR are approximately the same for both phantoms. But for SSIM the cross sections phantom yields a non–consistent result: SSIM measures the structural similarity, but this kind of phantom data shows practically no structures.

This experiment highlights that the use of inappropriate phantoms may drive to erroneous conclusions in the assessment of filtering performance. Besides, it is useful to show the convenience of the phantom data design presented in Section 5.4.1. Only this data set yields consistent results for both performance measures (since it contains realistic image structures such as edges, changes in contrast, and so on). Furthermore, the cross sections phantom may yield clearly incorrect conclusions: the lack of the aforementioned image structures bias the comparison towards those filters which remove a higher amount of noise at the expense of blurring the edges, as is the case with CA.

5.4.5 Bias removal

The deficiencies of some synthetic phantoms often used in the literature have been evidenced in the previous Section. However, these phantoms may be useful to study some features of the filters such as the ability to remove the bias in Rician distributed signals. A diffusion tensor has been fitted with the WLS technique described in [Sal05] for each voxel of the (filtered) phantom in Fig. 5.4, and the FA and MD have been computed. LMMSE–5, LMMSE–15, and UNLM–1 are compared in Fig. 5.6, where 2-D histogram plots in the FA/MD plane are shown.

For the original data, the estimation is clearly biased, and yet it is impossible to distinguish between the two clusters on the left–hand side. Moreover, FA values greater than 1 and MD values minor than 0 may be found due to the presence of negative eigenvalues of the tensor. Once the data is filtered, all the clusters of the histograms are centered in their corresponding ground–truth centroids, which is, the bias introduced by Rician noise is removed. Comparing LMMSE–5 and LMMSE–15, the use of a larger number of gradient directions improves the filtering, especially for the cluster on the left–hand side. UNLM–1 yields the best results, showing practically no outliers far from the ground–truth centroids. However, as shown in the previous Section, the goodness of each filter cannot be inferred from this experiment, which is only intended to show that in fact all the presented filters are able to remove the bias of Rician signals.

5.4.6 *In vivo* experiments

In the previous Sections, the noise power was assumed to be known. However, when dealing with real data, σ_n has to be estimated. Although this problem is

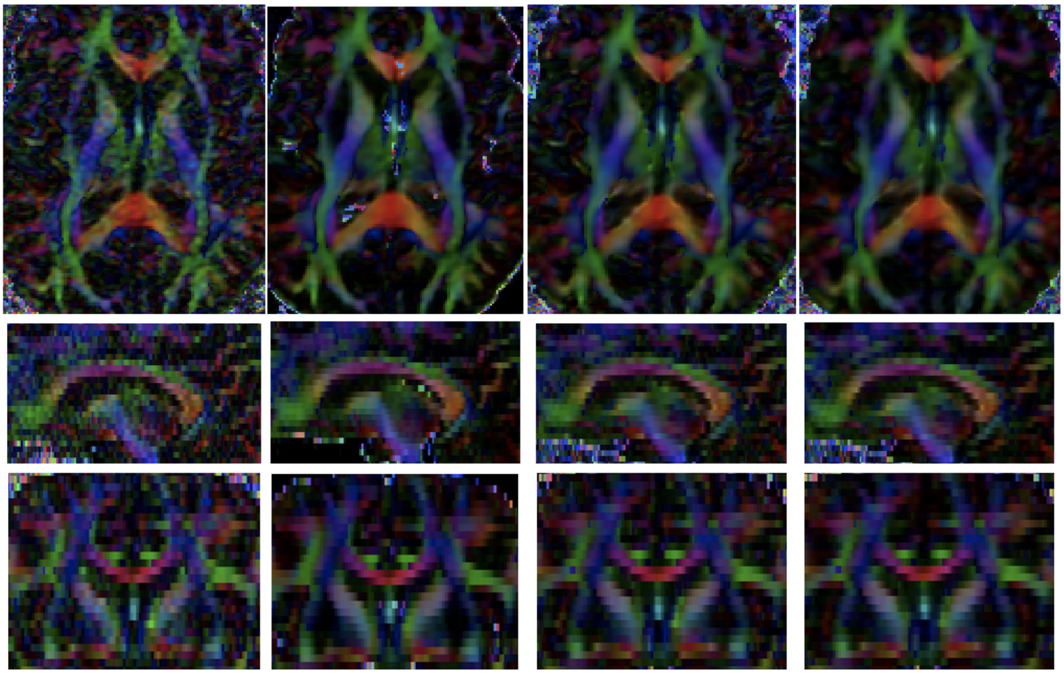


Figure 5.7: Color by orientation for CDR1. From left to right: original, UNLM–5, LMMSE–1 and LMMSE–15. Top: axial; middle: sagital; bottom: coronal.

beyond the scope of this dissertation, Appendix C shows a brief review of the techniques commonly used for this task. In what follows, one of the methods proposed in [AF09] is used: the background of the baseline image is automatically segmented [Ots79], and at each voxel the local first order moment is computed. The mode of the distribution of this moment, computed as the position of the maximum in its histogram, is corrected with a factor $\sqrt{2/\pi}$ to yield the estimated value of σ_n. This value, together with the peak amplitudes $\mathscr{A}$ of the baseline, the RMSE averaged for all gradient images, the PSNR, and the SNR, are shown in Table 5.2 for the real data sets used in the experiments of this Section.

In the first experiment, UNLM–5, LMMSE–1 and LMMSE–15 are compared over CDR1, as an example of conventional medium–high SNR volume with few gradients. Since the slice thickness is very coarse for this volume, a neighborhood of size $5 \times 5 \times 1$ is used for LMMSE; for UNLM, this is not an issue. Fig. 5.7 shows the result of color by orientation (see Section 2.6.1) after tensor estimation with WLS.

All methods are able to notably improve the coherence in the orientation and remove the noise, while adequately preserving the borders of the images. Comparing LMMSE–1 and LMMSE–15, the results are very similar: the use of joint information is not so critical in this case. Comparing LMMSE and UNLM, the latter produces a slightly greater blurring of the corpus callosum, whose structural information both LMMSE–1 and LMMSE–15 are able to preserve. On the contrary, in the union of the corticopontine tract and the inferior longitudinal fasciculus

	CDR1	BWH1	BWH2	CSIRO1
σ_n	38	65	65	10.5
Peak value($\mathscr{A}$)	5,161	25,000	25,000	2,500
RMSE	115	385	385	27
PSNR (dB)	42.7	51.1	51.7	47.54
SNR (dB)	9.62	15.45	15.45	8.20

Table 5.2: Noise power σ_n, peak value $\mathscr{A}$ in the baseline, RMSE, PSNR and SNR for some data sets.

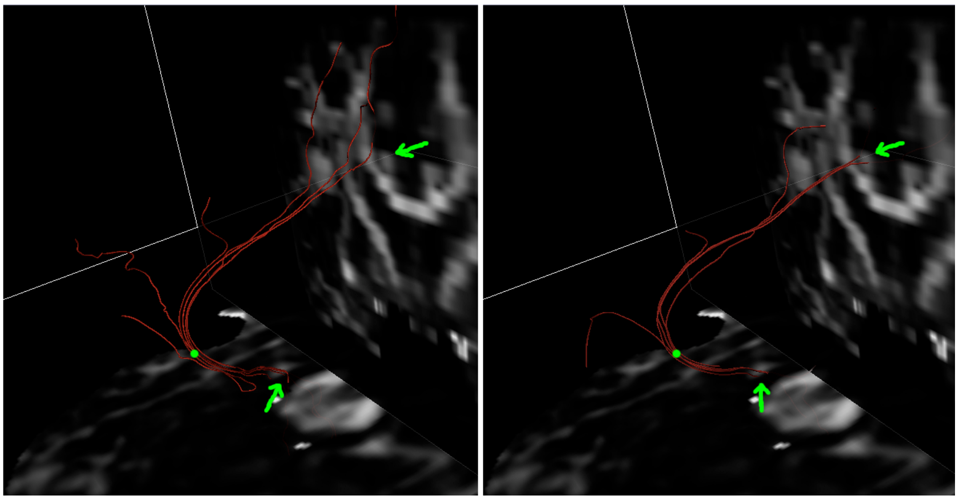


Figure 5.8: Fiber tracking of the cingulum for the original noisy data set (left) and the LMMSE–15 filtered data set (right). The green spots represent the seeding points for the tracking algorithm, and the green arrows indicate the exact points in the FA map corresponding to the cingulum in each slice [Mor05].

(the small structures at the right, bottom part of the axial view) UNLM–5 is able to better preserve the edges without blurring them. UNLM–N is more effective when the contrast between the structures and their surroundings is high enough, like in these small structures surrounded by a black background. In the corpus callosum, the contrast is not so high, and LMMSE performs better. When the contrast is high, the weights in UNLM are much greater for similar structures than they are for different anatomies. But for low contrast, it is impossible to distinguish the structures only by correlating similar patches; in consequence, all weights are quite similar, and UNLM accomplishes a rough average of voxels in a large neighborhood. The potential of UNLM relies on the ability to adequately adjust the non–local weights to distinguish between different structures.

To stress the importance of unbiased filtering in DTI, Fig. 5.8 shows an example of fiber bundles computed using tractography over the original noisy data set, and over the data set filtered with LMMSE–15. From the seeding points, a Runge–Kutta method of order 4 is used to track the direction of the cingulum, following the direction of the principal eigenvector at each point [Bas00b, Ten02]. The filtered volume yields anatomically correct results, and the reconstructed cingulum crosses the axial and the coronal slices at the correct locations [Mor05]. The noisy volume causes great errors in the location of fiber bundles, and what is more important, the tracking algorithm is not able to follow the cingulum in the upper brain, which is mixed–up with the corpus callosum.

In the next experiment, the capability of LMMSE–N in eq. (5.19) to overcome the drawbacks of joint LMMSE in eq. (5.14) is tested. Several configurations of LMMSE–N are used to filter BWH1, with a neighborhood size of $5 \times 5 \times 3$. For comparison purposes, UNLM–5 is used in Fig 5.9. The difference between LMMSE–1 and LMMSE–N is now very clear: LMMSE–1 is not able to properly remove the noise, and the structural information of the image is mostly lost; with LMMSE–N, the noise is removed and not only the structures are preserved, but they are even enhanced (see for example the left–bottom part of the slice, red circle). Comparing LMMSE–15 and LMMSE–51, no noticeable differences may be found. But comparing to joint LMMSE, the details with LMMSE–N are better preserved, the over–blurring is avoided, and moreover the ringing artifacts near the ventricles and in the outer contours of the brain mostly disappear. It is worth noticing that the artifacts are avoided even for LMMSE–N with $N = Z_g = 51$, so it follows that

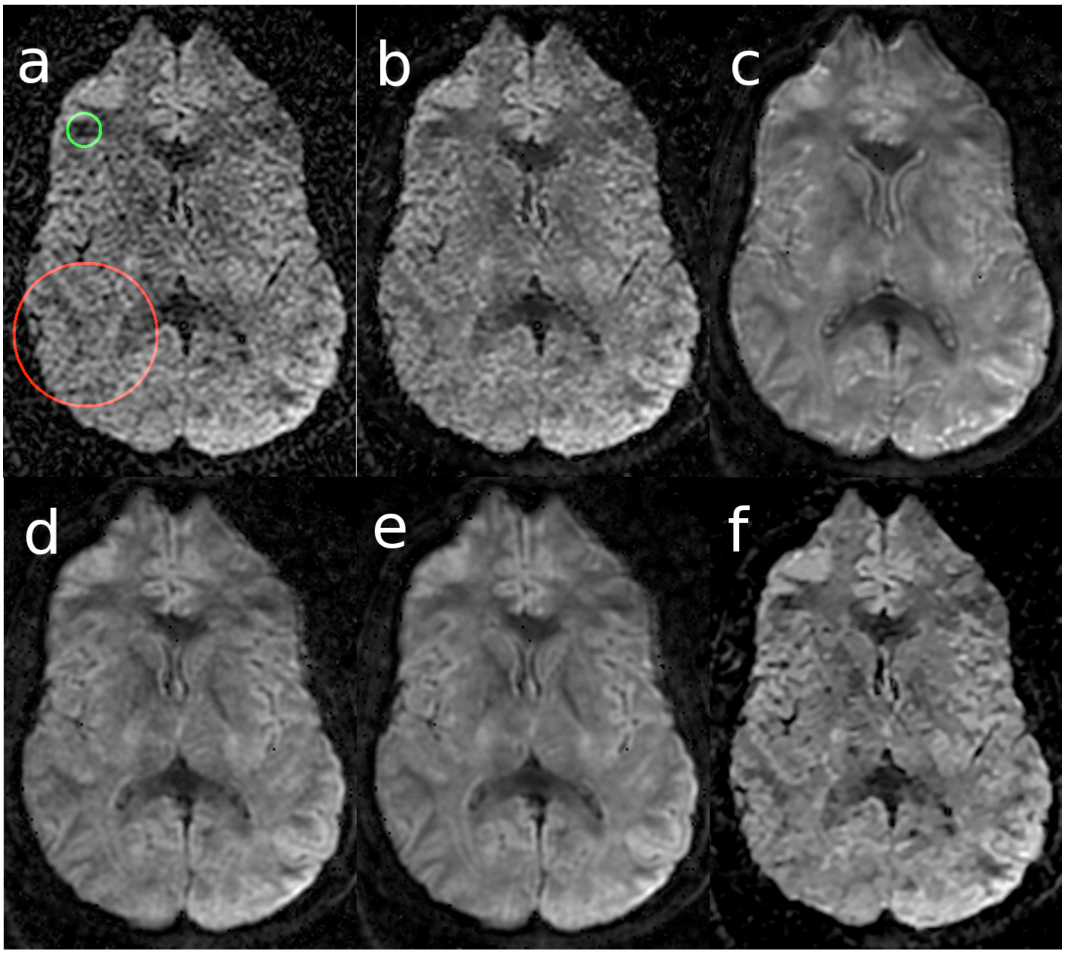


Figure 5.9: Central slice of the first gradient direction ($\mathbf{g}_i = [1,0,0]^T$) of BWH1: (a) original noisy volume, (b) LMMSE–1, (c) joint LMMSE (see [TV08]), (d) LMMSE–15, (e) LMMSE-51 and (f) UNLM–5.

the main source of error in the model of [TV08] is the joint estimation of both the baseline images and the DWI gradients, and not the eddy current distortion. Comparing now to UNLM, this filter is able to better remove the noise that LMMSE–1, while adequately preserving the edges. But the convolved structures in the peripheral brain are almost hindered, and only LMMSE–*N* is able to reveal them. Once again, LMMSE–*N* is able to properly enhance low–contrast structures, while UNLM is preferable to preserve the edges when the contrast is higher (see especially the green circle in the top–left part of the image, for which LMMSE seems to completely blur the structure).

The previous result is not completely conclusive, since only one gradient direction has been shown. This clearly does not suffice to assess the capability of the filters to recover the rich anatomical information DWI data sets provide. Given the large number of gradient directions in BWH1, it may be analyzed by means of HARDI techniques. In particular, the Generalized Anisotropy (GA) at each voxel may be computed [Öza05,Des06] as described in eq. (2.42). Fig. 5.10 shows the GA for UNLM–5, LMMSE–1 and LMMSE–15. All filters are able to improve the coherence on the anatomical structures while reducing the noise. However, LMMSE–15 is able to yield results more correct anatomically. This is especially noticeable in the middle (slice 29) and upper (slice 48) brain, although the cerebellar peduncle is also better delimited with LMMSE–15. In slice 48, LMMSE–15 is able to better resolve a large structure such as the corpus callosum, but also to recover the small nervous terminations in the peripheral brain. The advantage of LMMSE–*N* when

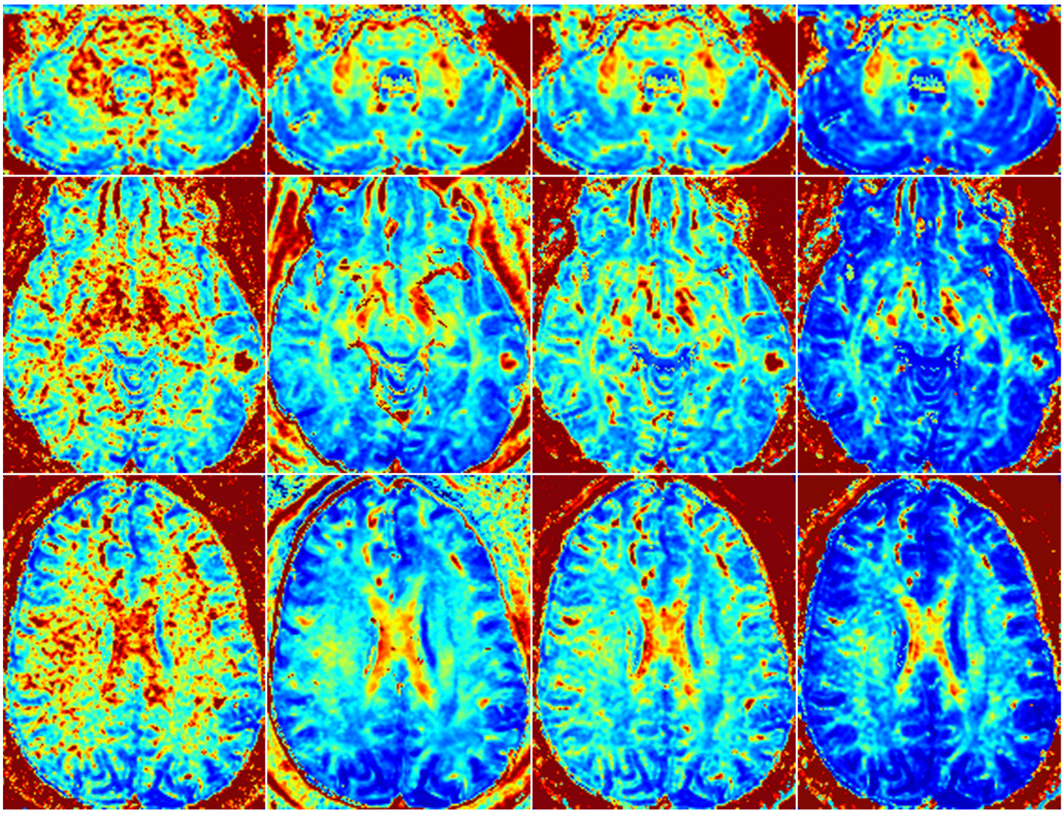


Figure 5.10: Pseudo–color representations of the GA for axial slices 15 (up), 29 (middle), and 48 (bottom), of a total of 78 slices from BWH1: original image (first column), UNLM–5 (second column), LMMSE–1 (third column) and LMMSE–15 (fourth column).

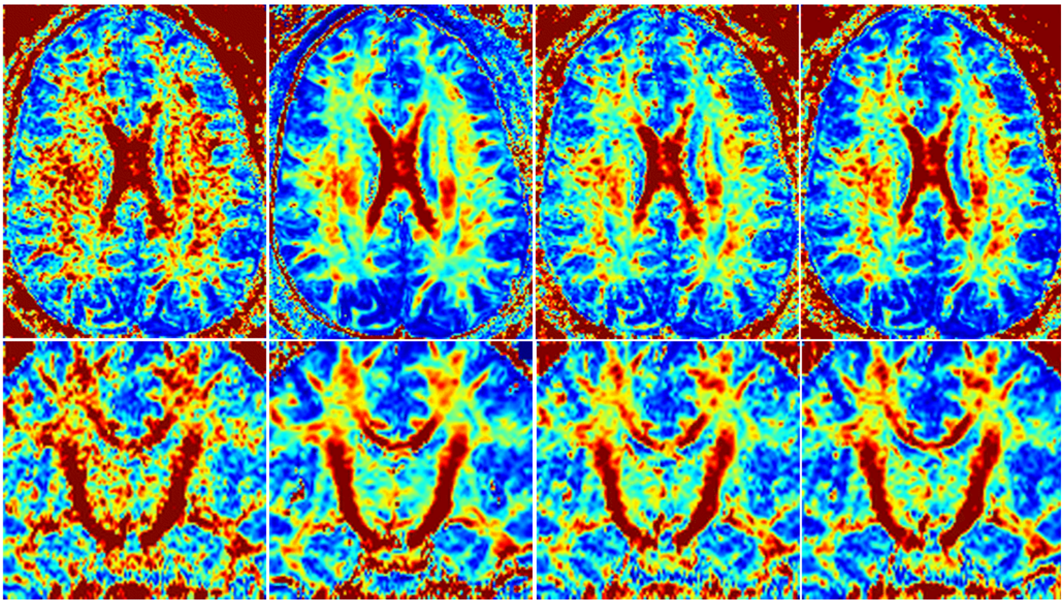


Figure 5.11: Pseudo–color representations of the FA for an axial (top) and a coronal (bottom) view of BWH1: original image (first column), UNLM–5 (second column), LMMSE–1 (third column) and LMMSE–15 (fourth column).

compared to LMMSE–1 is obvious in this case (see the third and fourth columns of Fig. 5.10). Besides, LMMSE–15 does not produce neither over–blurring (it is even able to better preserve the edges of the GA image) nor ringing artifacts, overcoming

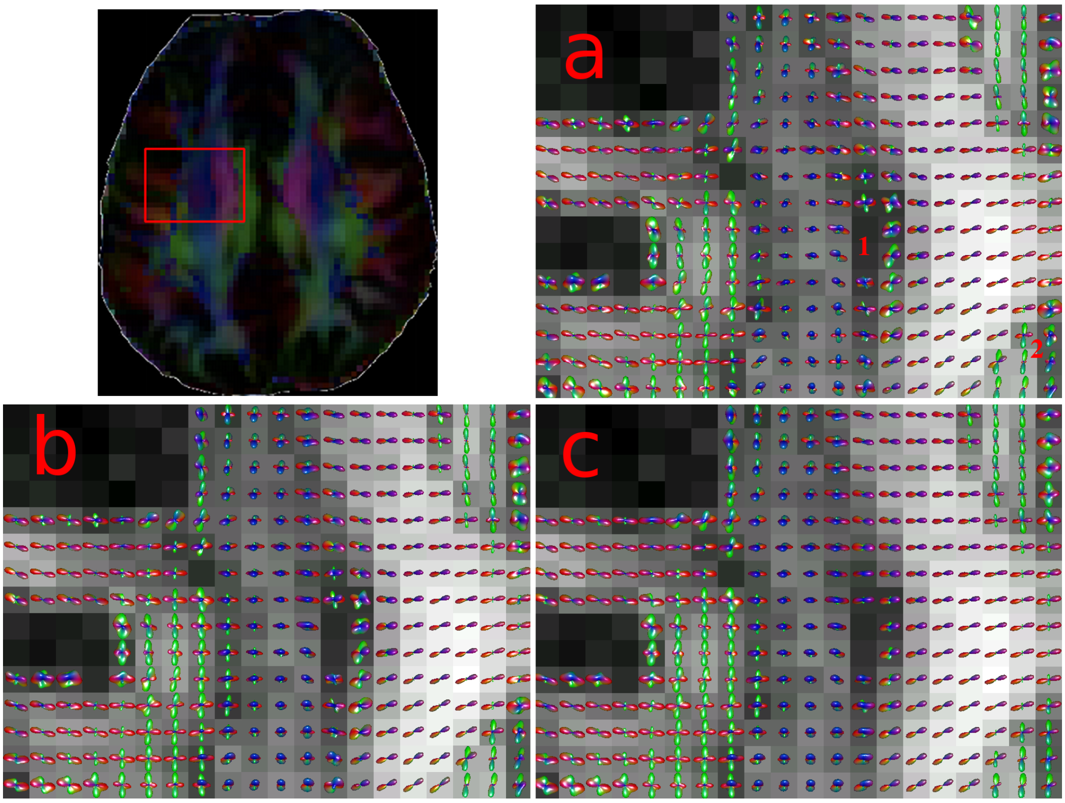


Figure 5.12: Glyph representations of fiber orientation functions obtained with the DOT for: (a) the original data set, (b) LMMSE–1 and (c) LMMSE–60. Several tracts of interest may be identified: the superior longitudinal fasciculus (green glyphs on the left–hand side); the superior corona radiata (blue glyphs); the cingulus (green glyphs on the right–hand side) and the corpus callosum (red glyphs on the right–hand side).

the main limitations of the preliminary approach in [TV08].

For the sake of completeness, Fig. 5.11 shows FA maps (over tensor fields reconstructed by WLS) for the same algorithms tested in Fig. 5.10. In this case the differences in the performance are not so important. Yet, LMMSE–15 is able to better remove the noise and yield a cleaner image than LMMSE–1 does, while correctly preserving the edges of the image. Comparing UNLM–5 and LMMSE–15 in the axial view, the results are slightly better for the former. UNLM–5 yields a cleaner image without blurring the edges.The difference is especially obtrusive at the small structures in the peripheral brain. In the coronal view, due to the smaller resolution in the z axis direction, the qualitative performance of all filters looks similar, although UNLM–5 seems to blur the structures in the top of the brain.

The last experiment is intended to stress the advantages of accounting for joint information in HARDI data sets. To that end, a typical volume for this kind of techniques, CSIRO1, has been chosen. An axial slice in the upper brain, where a number of important fiber bundles are known to coexist [Mor05], has been taken. At each voxel, the Diffusion Orientation Transform (DOT) proposed in [Öza06] has been used to infer the principal fiber orientations, and a glyph representing this information has been represented (see Section 2.6.1)[6] in Fig. 5.12. LMMSE–60 is

---

[6] The estimation of complex fiber orientations configurations is the main focus of this thesis, so the remaining Chapters are devoted to this topic. In particular, the kind of information to represent, the techniques used to infer it, and its visual representation will be thoroughly analyzed. At this point, it is enough to consider the glyphs as entities representing information about the orientation of fiber bundles in an intuitive way. This interpretation is enough to comprehend the impact of the filtering

N	1	5	10	20	40	60 (All)
LMMSE–N	25	60	115	325	1,130	75
UNLM–N	6,300	*31,500*	*63,000*	*126,000*	*252,000*	*378,000*

Table 5.3: Execution times (seconds) of LMMSE–N and UNLM–N for different values of N.

able to yield an overall smoother result, but the fiber crossings are not hindered at all by the filtering. To stress the different performance of LMMSE–1 and LMMSE–60, two regions of interest have been marked in the image. In (1), the DOT is not able to resolve the crossings between the corona radiata and the corpus callosum. The image filtered with LMMSE–1 does not improve the estimation, but LMMSE–60 is able to amend the coherence of the orientation information in this region and account for the fiber crossings. The situation is even clearer in (2), where only LMMSE–60 is able to improve the estimation and correctly resolve both the corpus callosum and the cingulus (this comment holds for all glyphs in the last column of the images, both in the top and the bottom regions).

5.4.7 Some notes on computational complexity

Since the main drawback of UNLM is known to be its great computational load, it is worth to include a comparison regarding this issue. CSIRO1 is used as a benchmark to run the filters programed with the ITK libraries [Iba05] in a 16-core, 32 GB RAM machine in a Linux environment. The execution times (ET) provided are measured with the `time` command, showing here the total *user* time (total time for all threads). Since both algorithms have been programmed in a full multi–threaded way, the comparison is fair. The I/O time has been subtracted from the computation in all cases. It is easy to comprehend that the ET of UNLM–N scales linearly with the number of voxels in the neighborhood $\mathcal{N}_x'$ and with N. From the work by [Man08] it is known that using a smaller $\mathcal{N}_x'$ than $7\times7\times3$ notably worsens the results, so this size has been kept as a standard. The ET has been measured only for UNLM–1, and this result has been extrapolated to UNLM–N. For LMMSE–N, the neighborhood is of size $5\times5\times3$. The dependency with N is not so clear in this case, so the ET is measured for a wide range of values of N. Results are shown in Table 5.3.

LMMSE is several orders of magnitude faster than UNLM in all cases. Some optimizations have been proposed in [Cou08] to speed–up UNLM, mainly a voxel pre–selection and a block–wise implementation. Even when these two techniques may reduce the ET roughly 30 to 50 times, LMMSE is still faster; for example, if for UNLM–1 the best case is considered, i.e. the ET is reduced 50 times, the ET is 126 seconds, slower than LMMSE-10. Besides, the block–wise implementation (which supposes the main acceleration of the algorithm) of UNLM worsens the filtering accuracy, as reported in [Cou08]. Yet, it is not clear how the block–wise technique could be extended to UNLM–N. As a final comment, if $N = Z_g$, LMMSE–N is only 3 times slower than the conventional LMMSE.

5.4.8 Discussion

The unbiased filtering of DWI data sets not only improves the visual results in DTI or HARDI, but also meliorates the anatomical correctness of the quantitative analysis. The most evident example presented is in Fig. 5.8, where the pre–processing

---

approaches in this experiment.

allows to better track the fiber bundles of interest, but the advantage is clear as well in the FA/GA maps and glyph representations.

Regarding the use of joint information, it has been shown that it may be advantageous in many situations. DWI data sets show very particular characteristics regarding the underlying anatomical and diffusion structures of the white matter. Therefore, it seems logical to use *ad hoc* techniques instead of just filtering each component as if it was an independent data set. Although in conventional DTI the improvement is not so relevant (see Fig. 5.11), and yet in some situations of high SNR may be only marginal (see Fig. 5.7), it has been evidenced that for more general HARDI approaches it is very important to account for all the information in the data set simultaneously.

The estimation of the diffusion tensor by WLS introduces an extra regularization of data, since only six degrees of freedom need to be estimated from a typically larger number of gradient directions. For this reason, all filters perform in a similar way for DTI, whenever they are able to remove the bias of Rician noise (see Section 5.4.5). Nevertheless, it has been previously reported that the estimation of the diffusion tensor without previous filtering may introduce errors which cannot be recovered by the regularization inherent to WLS [Bas00a, Jon04]. In this sense, it has been shown in Section 5.4.6 that all filters notably improve the results with respect to the noisy data sets in all situations. It is worth notice that using joint information to process the data sets is useful in conventional DTI as well (although not so critical as in HARDI). It has been shown that LMMSE–N is able to improve the results even for small N, while the computation times are kept reasonable even for large N (this is especially the case for $N = Z_g$, see Table 5.3). Although the benefit is not so evident in the *in vivo* experiments, the quantitative evaluation presented in Sections 5.4.3 and 5.4.5 suggests that quantitative DTI may benefit from the joint information filtering approaches.

In HARDI techniques, the prediction of fiber orientations is often an ill–posed problem, so the regularization in the estimation of the orientation information is not so profitable. In these scenarios, the use of joint information becomes more relevant, and the filters accounting for it are able both to better remove the noise and preserve the anatomical structures (see Figs. 5.10 and 5.12).

With respect to the filters compared, LMMSE–N shows the best overall performance for HARDI techniques, Fig. 5.10 clearly illustrating this behavior, and has been evidenced to improve the anatomical correctness of the GA maps and the fiber tracking. Besides, the estimation of fiber orientations in LMMSE–N filtered volumes yields smoother estimations and better resolves fiber crossings in well–known anatomies (see Fig. 5.12). For DTI, the advantage of UNLM–N is clear. This filter is able to better preserve the edges of the images while appropriately removing the noise. Although it may produce a slight over–blurring of large, low–contrast structures in some situations (like the corpus callosum in the axial view of Fig. 5.7 or the top of the coronal view of Fig. 5.11), its ability to preserve and denoise small structures is a clear advantage when compared to LMMSE–N. However, LMMSE–N has a much lighter computational load (see Table 5.3) with results similar (although slightly worse) to UNLM–N, which makes it a very attractive alternative. Moreover, the advantage of the use of joint information in UNLM–N is only marginal (see Fig. 5.3), which is another weakness when compared to LMMSE–N. Generally speaking, UNLM–N is useful when the non–local weights applied to the averaged voxels can accurately discriminate between different structures of the images. Since these weights are found looking for similar patches in non–local neighborhoods, there are two main factors which affect the accuracy of this filter:

1. If the contrast (difference in the gray levels) between the structures to distinguish is low, the weights applied to the voxels in each structure will be similar, yielding an excessive blurring. This assertion is supported by the experiments in Figs. 5.7 and 5.9.
2. In low–SNR scenarios, the computation of distances in the space of gray–levels will be strongly affected by the noise, increasing the uncertainty in the values of the non–local weights. As as result, it will be more difficult to distinguish the structures of the image. The experiments in Section 5.4.3 evidence this behavior.

The previous discussion may help to choose a particular filtering technique in a given situation. The choice depends on several factors such as the SNR of the data set, the anatomical structures of interest, the scanning parameters, and of course the requirements on computation time.

5.5 Generalization to non–central Chi DWI data sets

To this point, all filtering techniques (and validation methodologies) have been centered on the Rician case. Without any doubt, this is the most important case nowadays, since all single–coil systems and the majority of pMRI protocols yield data sets described by this statistics. This is easy to notice reviewing the literature on DWI filtering: there are many works focused on Rician noise removal, but corresponding approaches to filter non–central Chi signals have not been proposed. Although the Rician case is currently of capital importance, it was stated in Chapter 3 that protocols driving to non–central Chi signals (such as multiple–coil MRI or GRAPPA reconstructed images) are gaining an increasing interest.

As stated in Section 5.2, most of the filtering techniques for Rician images are based on the even–order moments of this distribution. In particular, this is the case for LMMSE–N and UNLM–N. The even order moments of the non–central Chi distribution are equally easy to compute, so the extension of the methods presented to this probability model is straightforward. To begin with, the analogous results to eqs. (5.2) and (5.3) are:

$$\begin{aligned} \langle M_i^2 \rangle &= \langle A_{L,i}^2 \rangle + 2L\sigma_n^2; && (5.29)\\ \langle M_i^4 \rangle &= \langle A_{L,i}^4 \rangle + 4L(L+1)\sigma_n^2 \langle A_{L,i}^2 \rangle + 4L(L+1)\sigma_n^4, && (5.30) \end{aligned}$$

where, with the notation in Chapters 3 and 4, $A_{L,i}^2 = \sum_{l=1}^{L} A_{c,l,i}^2 + A_{s,l,i}^2$, for $A_{c,l,i} + jA_{s,l,i}$ the complex signal corresponding to the i–th gradient in coil l, and L the number of coils. For $L = 1$, these expressions particularize to those of eqs. (5.2) and (5.3).

5.5.1 LMMSE–N for non–central Chi signals

The expression of the LMMSE estimator only depends on the first and second order moments of data, so eq. (5.14) holds without changes (in this case the estimator is for $\mathbf{A}_L^2$ and not $\mathbf{A}^2$):

$$\widehat{\mathbf{A}_L^2} = \langle \mathbf{A}_L^2 \rangle + \mathbf{C}_{A_L^2 M^2} \mathbf{C}_{M^2 M^2}^{-1} \left(\mathbf{M}^2 - \langle \mathbf{M}^2 \rangle \right). \tag{5.31}$$

The estimation of $\langle \mathbf{M}^2 \rangle$ is performed as a local average, and $\langle \mathbf{A}_L^2 \rangle$ computed from eq. (5.29). For the covariance matrices $\mathbf{C}_{A_L^2 M^2}$ and $\mathbf{C}_{M^2 M^2}$, a similar analysis as the one carried out in Appendix B yields, whenever the noise may be considered uncorrelated between different coils:

$$\begin{aligned} \mathbf{C}_{A_L^2 M^2} &= \varsigma_L \langle \mathbf{A}_L^2 \rangle \langle \mathbf{A}_L^2 \rangle^T; \\ \mathbf{C}_{M^2 M^2} &= \varsigma_L \langle \mathbf{A}_L^2 \rangle \langle \mathbf{A}_L^2 \rangle^T + 4\sigma_n^2 \mathrm{diag}\left(\langle \mathbf{A}_L^2 \rangle \right) + 4L\sigma_n^4 \mathrm{I}_Z, \end{aligned} \tag{5.32}$$

for:

$$\varsigma_L = \frac{\langle A_{L,\beta}^4 \rangle - \langle A_{L,\beta}^2 \rangle^2}{\langle A_{L,\beta}^2 \rangle^2}.$$

The recursive rule to approximate the inverse of $\mathbf{C}_{M^2M^2}$ is now valid if $A_{L,i}^2 > L\sigma_n^2$ for all gradient images i, and reads:

$$\begin{aligned} \mathbf{C}_{M^2M^2}^{-1} \left(\mathbf{M}^2 - \langle \mathbf{M}^2 \rangle \right) &\simeq \mathbf{w}^{\mathscr{O}}; \\ \mathbf{w}^{l+1} &= \left(\mathbf{M}^2 - \langle \mathbf{M}^2 \rangle \right) - 4L\sigma_n^2 \widetilde{\mathbf{C}}_{M^2M^2}^{-1} \mathbf{w}^l; \\ \mathbf{w}^0 &= \left(\mathbf{M}^2 - \langle \mathbf{M}^2 \rangle \right); \\ \widetilde{\mathbf{C}}_{M^2M^2}^{-1} &= \eta_L \mathbf{1}\mathbf{1}^T + \mathrm{diag}(\mathbf{e}_L), \end{aligned} \tag{5.33}$$

for:

$$\eta_L = -\left(4\sigma_n^2 \left(\frac{4\sigma_n^2}{\varsigma_L} + \sum_{i=0}^{Z} \langle A_{L,i}^2 \rangle \right) \right)^{-1}; \ \mathbf{e}_{L,i} = \left(4\sigma_n^2 \langle A_{L,i}^2 \rangle \right)^{-1}.$$

Obviously, the extension of LMMSE–N is straightforward for non–central Chi signals, since this filter works by applying the method above described for a subset of gradient images, which are all of them non–central Chi distributed.

5.5.2 UNLM–N for non–central Chi signals

The extension of UNLM-N is even simpler. Eq (5.20) may be directly extrapolated changing the bias removal term to take into account the result in eq. (5.29):

$$\widehat{A_{L,i}^2}(\mathbf{x}) = \sum_{j \in \Theta_i^N} \sum_{\mathbf{c} \in \mathscr{N}_x'} w_i^j(\mathbf{x}, \mathbf{c}) M_j^2(\mathbf{c}) - 2L\sigma_n^2, \tag{5.34}$$

where the weights w_i^j are computed exactly the same as in eq. (5.21).

5.6 Conclusion

In this Chapter, empirical evidences of the theoretical study carried out in Chapters 3 and 4 have been given: the noise in DWI data sets induces large errors which cannot be recovered by further regularization. The experiment in Fig. 5.8 is especially enlightening in this sense: even for a high SNR scenario, the noise hinders the actual anatomy of the cingulum, despite of the regularization introduced by the Runge–Kutta integration method. Once again, the need for an adequate denoising scheme is evident.

Although other approaches, such as tensor estimation accounting for the Rician model, are possible, the prior filtering of DWI data sets has been demonstrated highly valuable. Going back to the experiment in Fig. 5.8, this pre–processing allows to correctly resolve the fiber architecture, yielding anatomically correct results. The improvement achieved by this methodology is also evident in all the experiments presented in this Chapter.

The advantage of filtering the DWI prior to the estimation of diffusion propagators relies on the independence of this process from the estimation itself. Once the volume has been denoised, the diffusion propagator may be calculated assuming that the signal in the **q**–space is almost noise–free, so the noise model may be obviated. Conversely, the denoising algorithm has not to take into account the particular technique which is to be used for fiber orientation estimation. Nevertheless, it has been shown that each filtering technique is better suited for each given application (for example, UNLM–N is preferable for conventional DTI, but LMMSE–N is more accurate for general HARDI techniques).

Another important conclusion is the need to take into account the particular characteristics of DWI data sets, previously suggested in [Fil07]. Gradient images in the DWI volume are not isolated, independent channels in a multi–channel image, but instead correspond to single coordinates in the 6D space of the Diffusion Imaging framework (see Fig. 2.6). As such, filtering each image separately is not the most adequate method.

One important issue is that the filtering techniques described are based on the computation of local (LMMSE–N) or non–local (UNLM–N) averages of voxels. This computation makes sense as long as the statistics of the voxels averaged are the same. The analysis carried out in Chapter 3 suggests that this is not the case for pMRI techniques. Fortunately, this same analysis shows that the power of noise varies slowly enough (in fact, much slower than the variation of the structures of interest in the image) to consider that the statistics of the voxels averaged are very similar. It is so even for UNLM–N, since the non–local neighborhoods used for this filter are yet small compared to the range of variation of the power of noise. Nevertheless, this is a theoretical limit for the use of arbitrarily large neighborhoods with UNLM–N. For both filters (and any other needing the power of noise as an input), variability maps of σ_n, similar to those provided in Chapter 3 for GRAPPA reconstructions, could be used to refine the model.

The final remark in this discussion is centered on the generalization of the filtering techniques to the non–central Chi case. Although the extension from Rician–adapted methods is straightforward, the filters proposed are the first approach to of denoising non–central Chi images to the knowledge of the author. These results, together with the analysis in Chapter 3 which warranties the slow variability of the noise power for GRAPPA reconstructions, allow to take advantage of all the concepts presented in this Chapter for this important imaging protocol.

6

Probabilistic estimation of fiber orientations in the white matter: the Orientation Probability Density Transform

This Chapter, together with Chapter 7, contains the main contribution of this thesis. It is focused on the development of a novel estimator of the directions of fiber bundles in the white matter when the Gaussian assumption of DTI does not hold. The first part is dedicated to briefly survey the theoretical foundations of HARDI to analyze the requirements that the estimator has to meet. Three concepts are detailed later on: the Orientation Probability Density Function, the Funk–Radon transform, and the expansion of functions in the basis of Spherical Harmonics. The combined use of these three techniques is the key for the development of an estimator fulfilling the aforementioned requirements, and is the basis for the definition of the Orientation Probability Density Transform (OPDT). The description of the OPDT is the main topic in this Chapter. After the theoretical analysis, the properties of the novel estimator are empirically tested in a wide variety of numerical examples, comparing it to other related approaches. The OPDT shows important advantages in a number of situations, especially for low values of the weighting parameter b_0, but also an important limitation, due to the use of the Funk–Radon transform, which will focus the next Chapter.

6.1 Introduction

It was stated in Chapter 2 that the ultimate aim in Diffusion Imaging is the estimation of the directions of fiber bundles, which are in turn associated to the probability of diffusion of water molecules in these same directions. Although the assumption of a Gaussian propagator made in DTI is useful for some tasks, this approach fails to describe complex neural architectures such as fiber crossing, bending or kissing (see Section 2.7) In these cases, the relationship given by eq. (2.27) has general validity whenever the short–pulses condition ($\delta \ll \Delta$) holds. Eq. (2.27) is reproduced here for convenience:

$$P(\mathbf{R}) = \mathfrak{F}\left\{E(\mathbf{q})\right\}(\mathbf{R}) = \iiint_{\mathbb{R}^3} E(\mathbf{q})\exp(-2\pi i\mathbf{q}\cdot\mathbf{R})d\mathbf{q}; \tag{6.1}$$

note that the dependence on the diffusion time Δ has been dropped down. This notation will be used throughout this Chapter. On the other hand, the modulus of the diffusion signal, $E(\mathbf{q})$, is always less than 1, so the general representation in eq. (2.39), reproduced next, is always possible:

$$E(\mathbf{q}) = \exp\left(-bD(\mathbf{q})\right), \tag{6.2}$$

where D is the positive Apparent Diffusion Coefficient (ADC), which generalizes the positive definite diffusion tensor $\mathscr{D}$. From eq. (6.2) it follows that the characterization of the ADC is equivalent to the characterization of $E(\mathbf{q})$. Therefore, sampling the value of either $E(\mathbf{q})$ or $D(\mathbf{q})$ for the whole **q**–space allows the explicit evaluation of the integral in eq. (6.1) to estimate the diffusion propagator. The marginal PDF for all possible directions $\mathbf{r} = \mathbf{R}/\|\mathbf{R}\|$ can be further computed to estimate the probability of the existence of a fiber bundle along $\mathbf{r}$. This is the principle of DSI. To reduce the great computational load required to sample the whole **q**–space, HARDI techniques perform a dense sampling of a subset of the **q**–space comprising all the spatial directions of a sphere of radius $\|\mathbf{q}\| = q_0$. All these concepts, together with a survey of the existing HARDI techniques, were reviewed in Sections 2.5 to 2.7, and the reader is referred there for a detailed analysis.

The main limitation of HARDI is that in general the ADC depends on both the direction $\mathbf{g}$ and the modulus q of $\mathbf{q}$, so sampling $E(\mathbf{q})$ in the sphere $\mathscr{S} = \{\mathbf{q} | \|\mathbf{q}\| = q_0\}$ gives only **incomplete information** on the behavior of the diffusion signal. HARDI techniques previously used in the literature deal with this problem in several ways (see Section 2.7 for details). In general, these solutions share one or more of the following drawbacks:

1. The lack of a true probabilistic interpretation of the estimates. Instead of the marginal PDF associated to the diffusion propagator, some other orientation information may be estimated **indirectly** (i.e. not through the marginalization of $P(\mathbf{R})$) from $E(\mathbf{q})$. This way, the need to characterize the entire diffusion signal may be avoided. Persistent Angular Structures [Jan03] and Q–Balls [Tuc03, Tuc04] are grounded on this principle. The DOT [Öza06] is not based on the same idea, but also shares this deficiency.
2. The need to make unrealistic assumptions on $E(\mathbf{q})$. If the behavior of $E(\mathbf{q})$ is extrapolated from the sampled value at q_0 for all possible $\mathbf{q}$, the whole **q**–space may be characterized in terms of the sphere $\mathscr{S}$. This way, the diffusion propagator may be explicitly computed using eq. (6.1), and $P(\mathbf{R})$ can be marginalized to obtain the desired probabilistic orientation information. The most common assumption, present in higher order tensors [Des06] and the Diffusion Orientation Transform (DOT) [Öza06], is that the ADC is constant for all q ($D(\mathbf{q}) \equiv D(\mathbf{g})$), which is quite unrealistic.
3. The use of parametric models for the attenuation signal or the diffusion propagator. In multi–tensor approaches [Kre05, Pel06], it is assumed that the diffusion propagator is a mixture of a finite number of Gaussians, which has to be known beforehand. Spherical deconvolution [And05,Des09,Tou07,Tou08] generalizes this approach to an infinite number of Gaussians, but all of them are assumed to have the same particular shape (covariance matrix). Continuous mixtures of Gaussians [Jia07] do not assume any particular shape, but the covariance matrix has to follow a prior Wishart distribution, which is rather restrictive. The problem with all these techniques is that the estimates are highly prejudiced by poor fits of the parametric model to real data. Additionally, the need to estimate the parameters of such model can also be a pitfall for its application.

Additionally, the requirement of a high b–value (in the order of 3000 s/mm^2) is also present in most of the current techniques. The main focus of this thesis is to find adequate estimators able to overcome (at least partially) these three pitfalls at the same time. The next three Sections (6.2 to 6.4) are respectively centered on each of these topics. The concepts discussed are of capital importance to understand the estimation techniques introduced in Chapters 6 and 7, which comprise the most important work in this dissertation.

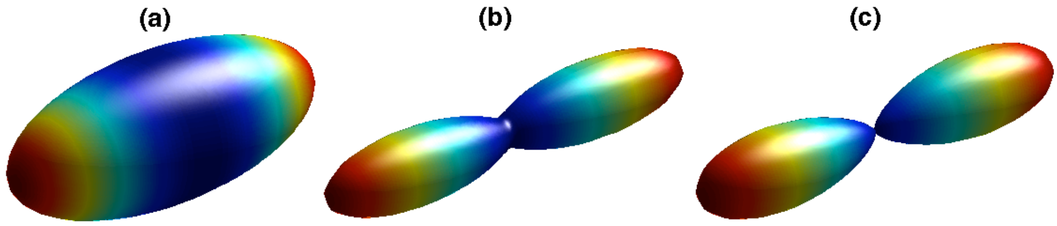


Figure 6.1: Examples of theoretically computed orientation functions for a simple tensor model with eigenvalues $[1.7, 0.3, 0.3] \cdot 10^{-3}\text{mm}^2/\text{s}$: (a) $\Psi(\mathbf{r}) = C_1 \cdot \left(\mathbf{r}^T \mathscr{D}^{-1}\mathbf{r}\right)^{-1/2}$; (b) $\Phi(\mathbf{r}) = C_2 \cdot \left(\mathbf{r}^T \mathscr{D}^{-1}\mathbf{r}\right)^{-3/2}$; (c) $\Upsilon(\mathbf{r}) = C_3 \cdot \exp\left(-R_0^2 \cdot \mathbf{r}^T \mathscr{D}^{-1}\mathbf{r}/4\tau\right)$ (for typical values $R_0 = 12\mu$m and $\tau = 20$ms). The value of the orientation function at each direction is represented as the distance of the surface to the origin.

6.2 Diffusion propagators and orientation information functions

Although $E(\mathbf{q})$ is in general not completely characterized in HARDI, it is not necessary either to completely characterize $P(\mathbf{R})$, but only its underlying orientation information: displacements in the same directions $\mathbf{r}$ are associated to the same fiber bundle. For each orientation $\mathbf{r}$ given by (θ, ϕ) in physics convention, the expression for the marginal probability density of a displacement along direction $\mathbf{r}$ is:

$$p(\mathbf{r}) \equiv p(\theta, \phi) = \int_0^\infty P(R\mathbf{r}) R^2 \sin\theta dR, \tag{6.3}$$

with $R = \|\mathbf{R}\|$ so $\mathbf{R} = R\mathbf{r}$. The term $R^2 \sin\theta$ is the Jacobian of the transformation to spherical coordinates, and it is therefore required to compute actual probabilities so that the integral of $P(R\mathbf{r})$ equals 1:

$$\begin{aligned}
\iiint_{\mathbb{R}^3} P(\mathbf{R}) d\mathbf{R} &= \int_0^\pi \int_0^{2\pi} \int_0^\infty P(R\mathbf{r}) R^2 \sin(\theta) dR d\phi d\theta \\
&= \int_0^\pi \int_0^{2\pi} \left(\int_0^\infty P(R\mathbf{r}) R^2 dR \right) \sin(\theta) d\phi d\theta \\
&= \int_0^\pi \int_0^{2\pi} \Phi(\theta, \phi) \sin(\theta) d\phi d\theta = 1,
\end{aligned} \tag{6.4}$$

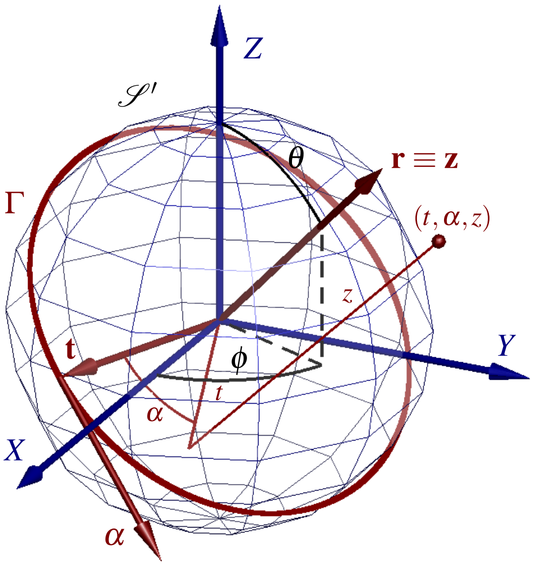
where $\Phi(\theta, \phi) \equiv \Phi(\mathbf{r})$ is the Orientation Probability Density Function (OPDF) defined as:

$$\Phi(\theta, \phi) \equiv \Phi(\mathbf{r}) = \int_0^\infty P(R\mathbf{r}) R^2 dR = \frac{1}{2} \int_{-\infty}^\infty P(R\mathbf{r}) R^2 dR, \tag{6.5}$$

since $P(\mathbf{R})$ has antipodal symmetry. In the last integral of eq. (6.4), the term $\sin(\theta)$ is once again due to the spherical coordinates system, i.e. it is required to compute the integral over the surface of the sphere $\mathscr{S}$. The orientation information for each direction $\mathbf{r}$ is characterized by the OPDF, which is a true probability density since its integral over all possible directions is 1. This definition has been already proposed before in [Wed05] as a “weighted radial summation”. Nevertheless, it turns out that it is usually easier to compute the Orientation Distribution Function (ODF) [Tuc03] as the **radial projection** of $P(\mathbf{R})$:

$$\Psi(\theta, \phi) \equiv \Psi(\mathbf{r}) = \int_0^\infty P(R\mathbf{r}) dR = \frac{1}{2} \int_{-\infty}^\infty P(R\mathbf{r}) dR. \tag{6.6}$$

The Jacobian of the spherical coordinates is dropped from the definition of $\Psi(\mathbf{r})$, so this function does not represent a true PDF [Tuc03, Tuc04]. In practice, this has the effect of blurring the orientation information. Other representations, such as the probability profile for a given radius R_0, $\Upsilon(\mathbf{r}) = P(R_0\mathbf{r})$, are possible as well [Öza06]. For Gaussian diffusion propagators, all these functions may be explicitly



Parametrization of Γ:

$$\Gamma \equiv \begin{cases} t = q_0 \\ \alpha = \alpha \quad , \; \alpha \in [0, 2\pi) \\ z = 0 \end{cases}$$

Correspondences between the cylindrical coordinates in the **q**–space and the dual **R**–space:

$$t \leftrightarrow u$$
$$\alpha \leftrightarrow \beta$$
$$z \leftrightarrow \vartheta$$

Figure 6.2: Auxiliary **cylindrical** coordinates system for the computation of the FRT. The equator $\Gamma \subset \mathscr{S}'$ is orthogonal to the direction of interest $\mathbf{r}$. It may be easily parametrized in the plane described by unit vectors $\mathbf{t}$ and $\boldsymbol{\alpha}$, so the integral of $E(q_0\mathbf{g})$ along Γ is trivial to compute if E is expressed in the new coordinates system. A similar system may be considered for the **R**–space, so the triplets (t, α, z) and (u, β, ϑ) are dual spaces with respect to the Fourier transform.

computed: Fig. 6.1 shows rendered surfaces (glyphs) of $\Psi(\mathbf{r})$, $\Phi(\mathbf{r})$ and $\Upsilon(\mathbf{r})$ for a typical tensor configuration: both $\Phi(\mathbf{r})$ and $\Upsilon(\mathbf{r})$ have sharper profiles than $\Psi(\mathbf{r})$. The blurring in the orientation information provided by the ODF ($\Psi(\mathbf{r})$) is a source of uncertainty in the estimation of fiber directions: two fiber bundles crossing in proximal directions would be more difficult to distinguish with the ODF than with the OPDF, since the two local maxima of the ODF could hinder each other. Besides, note that the OPDF is the only orientation function representing true probabilities.

6.3 The Funk–Radon transform

The advantage of using the ODF relies on the fact that it may be accurately approximated by the Funk–Radon transform (FRT) of the attenuation signal [Tuc03, Tuc04], $\mathscr{G}\{E(\mathbf{q})\}(\mathbf{r})$. The FRT is a linear integral transform defined on the space of functions whose domain is the sphere of radius 1. The input to the transform is a function taking values for each unitary direction $\mathbf{g}$, and its output is another function of the same kind, taking values in the transformed domain of unitary directions $\mathbf{r}$. For a given output direction $\mathbf{r}$, the FRT of $E(q_0\mathbf{g})$ is computed as the integral of the function along the equator perpendicular to the direction $\mathbf{r}$:

$$\mathscr{G}\left\{E(q_0\mathbf{g})\right\}(\mathbf{r}) \triangleq \iint_{\mathscr{S}'} \delta(\mathbf{r}^T\mathbf{g})E(q_0\mathbf{g})d\mathbf{g} = \frac{1}{q_0^2}\iint_{\mathscr{S}} \delta(\mathbf{r}^T\mathbf{q})E(\mathbf{q})d\mathbf{q}, \tag{6.7}$$

where $\mathscr{S}' = \{\mathbf{q} \,|\, \|\mathbf{q}\| = 1\}$. To understand the relation between the dual variables[1] $\mathbf{g}$ and $\mathbf{r}$, consider the cylindrical coordinates system (t, α, z) depicted in Fig. 6.2.

---

[1] The variables $\mathbf{g}$ and $\mathbf{r}$ are normalized (unit radius) versions of $\mathbf{q}$ and $\mathbf{R}$. The former is a dual pair with respect to the FRT, and the latter is a dual pair with respect to the Fourier transform. Instead of: 1.- Extrapolating the information from $q_0\mathbf{g}$ to the whole $\mathbf{q}$–space; 2.- Using the Fourier relation to estimate the diffusion propagator; 3.- Marginalizing $P(\mathbf{R})$ to compute the orientation information, the FRT directly relates the normalized domains of $\mathbf{g}$ and $\mathbf{r}$, so there is no need to know the behavior of $E(\mathbf{q})$ outside $\mathscr{S}$, i.e., outside of the domain of $\mathbf{g}$.

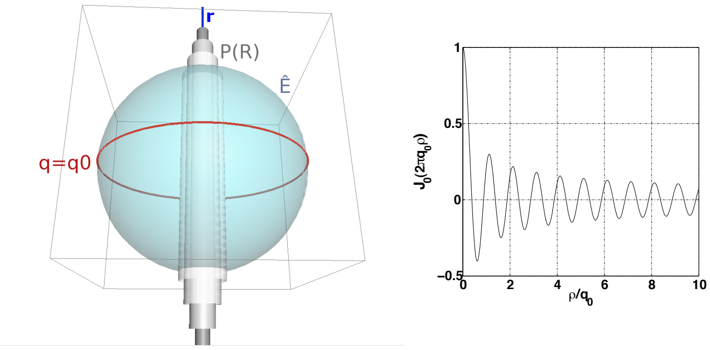


Figure 6.3: Illustration of the FRT. The integral of E (which has been sampled in the sphere) along the equator q_0 is equivalent to the integral of P inside the tubes along $\mathbf{r}$; each tube represent a local extreme of the Bessel function J_0 (right). The larger q_0, the narrower the tube.

The z axis is aligned with the direction of interest, $\mathbf{r}$, without loss of generality. The new coordinates system allows the computation of the FRT of $E(q_0\mathbf{g})$ for the unit direction $\mathbf{r}$ in a very easy way, since the equator under study admits a simple parametrization:

$$\mathscr{G}\left\{E(q_0\mathbf{g})\right\}(\mathbf{r}) = \frac{1}{q_0^2}\iint_{\mathscr{S}} \delta(\mathbf{r}^T\mathbf{q})E(\mathbf{q})d\mathbf{q} = \frac{1}{q_0^2}\int_0^{2\pi} \widetilde{E}(q_0,\alpha,0)q_0 d\alpha, \tag{6.8}$$

where $\widetilde{E}(t,\alpha,z)$ is the diffusion signal expressed in cylindrical coordinates. The diffusion propagator will have a different expression $\widetilde{P}(u,\beta,\vartheta)$ in the new system, but the Fourier relation given by eq. (6.1) still holds:

$$\begin{aligned}
\mathscr{G}\left\{E(q_0\mathbf{g})\right\}(\mathbf{r}) &= \frac{1}{q_0^2}\int_0^{2\pi} \widetilde{E}(q_0,\alpha,0)q_0 d\alpha \\
&= \frac{1}{q_0}\int_0^{2\pi}\left(\int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{2\pi}\widetilde{P}(u,\beta,\vartheta)\exp\left(j2\pi\left(q_0u\cos(\alpha-\beta)+0\vartheta\right)\right)u\,d\beta\,du\,d\vartheta\right)d\alpha \\
&= \frac{1}{q_0}\int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{2\pi}\widetilde{P}(u,\beta,\vartheta)\left(\int_0^{2\pi}\exp\left(j2\pi q_0u\cos(\alpha-\beta)\right)d\alpha\right)u\,d\beta\,du\,d\vartheta \\
&= \frac{2\pi}{q_0}\int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{2\pi}\widetilde{P}(u,\beta,\vartheta)J_0\left(2\pi q_0u\right)\,u\,d\beta\,du\,d\vartheta, \qquad (6.9)
\end{aligned}$$

where a well–known property of the Bessel function of the first kind J_0 has been exploited [Abr72]. This function is plotted in Fig. 6.3, right. Eq. (6.9) represents the integral of the diffusion propagator $\widetilde{P}(u,\beta,\vartheta)$ along the direction ϑ (i.e, the direction $\mathbf{r}$ of interest in the auxiliary coordinates system), pondered by the radial Bessel kernel J_0, which only depends on the transverse distance u. This computation is illustrated in Fig. 6.3, left. The line integral along $\mathbf{r}$ is replaced with a volumetric integral inside the tubes centered in the ϑ axis. Each one corresponds to a local extreme in the graph of J_0, so the tubes with minor radii contribute more to the integral. The FRT of $E(q_0\mathbf{g})$ constitutes an accurate estimate of the ODF as defined in eq. (6.6):

$$\begin{aligned}
\mathscr{G}\left\{E(q_0\mathbf{g})\right\}(\mathbf{r}) &= \frac{2\pi}{q_0}\int_0^{\infty}\int_0^{2\pi}J_0\left(2\pi q_0u\right)\left(\int_{-\infty}^{\infty}\widetilde{P}(u,\beta,\vartheta)d\vartheta\right)ud\beta du \\
&= \frac{4\pi}{q_0}\int_0^{\infty}\int_0^{2\pi}J_0\left(2\pi q_0u\right)\widetilde{\Psi}(u,\beta)ud\beta du \simeq Z(q_0)\Psi(\mathbf{r}), \qquad (6.10)
\end{aligned}$$

where $Z(q_0)$ is a constant (whose value may be neglected if the estimated ODF is normalized), and $\widetilde{\Psi}$ is the expression of Ψ in the cylindrical coordinates system. The integration along $\mathbf{r}$ is not exact: the blurring due to the Bessel kernel remains evident from eq. (6.10). Nevertheless, increasing the value of q_0 highly improves its accuracy, since the lobes of J_0 become more concentrated around $u = 0$.

Summarizing, the use of the FRT allows to **directly** relate the measured DWI in the domain of $\mathbf{g}$ (the sphere $\mathscr{S}$) with the orientation information in the domain of $\mathbf{r}$. Besides, note that absolutely any assumption on $E(\mathbf{q})$ outside $\mathscr{S}$ is required. The approximated estimation of the ODF requires **only** the sampling of $E(q_0\mathbf{g})$, which is completely characterized by HARDI data sets. This is the principle of Q–Ball imaging [Tuc03,Tuc04]. This way, the second requirement imposed in Section 6.1 is easily met at the expense of the following disadvantages:

1. The FRT, as conceived in Q–Ball imaging, is only useful to compute the ODF, which does not represent a true probabilistic information (see Section 6.2).
2. This computation is only approximated. For high values of q_0 the approximation becomes more accurate, so this problem may be somehow palliated. Note that the need for higher b–values than for conventional DTI was pointed out in Section 6.1 as a common issue in HARDI.

6.4 Spherical Harmonics Expansions

Both the DWI signal $E(q_0\mathbf{g})$ and the OPDF $\Phi(\mathbf{r})$ (or any other orientation information) are functions defined on the surface of a sphere, i.e., functions which take values for each unit direction $\mathbf{g}$ (or $\mathbf{r}$). For this reason it is very convenient to represent them in the basis of Spherical Harmonics (SH) [Fra02]. SH are defined as the eigenfunctions of the Laplace–Beltrami operator Δ_b. It is a differential operator representing the part of the Laplacian operator in the spherical coordinates system which does not depend on the radial coordinate; obviously, Δ_b applies for the kind of functions considered:

$$\Delta_b Y(\theta,\phi) = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y(\theta,\phi)}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 Y(\theta,\phi)}{\partial\phi^2} = \nu Y(\theta,\phi). \tag{6.11}$$

Eq. (6.11) represents a Sturm–Liouville problem which may be decoupled into two different sub–problems using separation of variables:

$$\begin{array}{rcl} Y(\theta,\phi) & = & A(\theta)B(\phi): \\ & & \left\{ \begin{array}{rcll} -m^2 & = & \dfrac{1}{B(\phi)}\dfrac{d^2B(\phi)}{d\phi^2} & \text{(SP1)} \\ m^2 & = & l(l+1)\sin^2(\theta) + \dfrac{\sin(\theta)}{A(\theta)}\dfrac{d}{d\theta}\left(\sin(\theta)\dfrac{dA(\theta)}{d\theta}\right) & \text{(SP2)} \end{array} \right. . \end{array} \tag{6.12}$$

The solution to SP1 is trivial: it is a complex exponential with wave number m. SP2 is the defining equation for the associated Legendre polynomials P_l^m [Abr72], so:

$$\begin{array}{rcl} A(\theta) & = & P_l^m(\cos\theta) \\ B(\phi) & = & e^{jm\phi} \\ Y_l^m(\theta,\phi) & = & \sqrt{\dfrac{2l+1}{4\pi}\dfrac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)\,e^{jm\phi} \end{array} \qquad l = 0\ldots\infty,\ m = -l\ldots l, \tag{6.13}$$

where l is the order of the SH, and the normalizing constant is chosen so that $Y_l^m(\theta,\phi)$ form an orthonormal set over the sphere of radius 1, $\mathscr{S}'$. In fact, from Sturm–Liouville theory it can be inferred that the (normalized) solutions to eq. (6.11) form an orthonormal basis for the functions defined on $\mathscr{S}'$. Casting eq. (6.13) into eq. (6.12) yields the actual value of the eigenvalues of the problem:

$$\Delta_b Y_l^m(\theta,\phi) = -l(l+1)Y_l^m(\theta,\phi),\ l=0\ldots\infty,\ m=-l\ldots l. \tag{6.14}$$

The basis functions $Y_l^m(\theta,\phi)$ are complex valued, but both the diffusion signal and the diffusion propagator are real (and positive) functions. Consequently, this basis is modified to represent the subspace of real valued functions as follows [Des07]:

$$\widetilde{Y}_l^m(\theta,\phi) = \begin{cases} \sqrt{2}\,\Re\left\{Y_l^m(\theta,\phi)\right\}, & -l \le m < 0 \\ Y_l^0(\theta,\phi), & m=0 \\ \sqrt{2}\,\Im\left\{Y_l^m(\theta,\phi)\right\}, & 0 < m \le l \end{cases}, \tag{6.15}$$

which is still an orthonormal basis with the same eigenvalues as in eq. (6.14). Finally, note that both the diffusion signal and the diffusion propagator have radial symmetry; it translates in that only even–order basis functions are required [Des07]. As a corollary, it follows that any real valued function $\mathscr{H}(\theta,\phi)$ defined on $\mathscr{S}'$, and radially symmetric, can be written in terms of its SH expansion:

$$\begin{aligned} \mathscr{H}(\theta,\phi) &= \sum_{l=0}^{\infty}\sum_{m=-2l}^{2l} C_{2l}^m\,\widetilde{Y}_{2l}^m(\theta,\phi). \\ C_{2l}^m &= \int_0^{\pi}\int_0^{2\pi} \mathscr{H}(\theta,\phi)\widetilde{Y}_{2l}^m(\theta,\phi)\sin(\theta)d\phi d\theta. \end{aligned} \tag{6.16}$$

The representation of orientation functions in terms of SH expansions is quite usual in HARDI [And05,Des06,Des07,Fra02,Öza06]. It has a number of important advantages:

1. It is non–parametric, so any real, radially symmetric function arbitrarily complex may be accurately approximated in terms of a truncation of the maximum order L of the basis functions.

2. Although the computation of C_{2l}^m in eq. (6.16) requires the computation of the integral defining the scalar product for each $2l$ and m, these coefficients can be computed very efficiently by means of LS fitting [Des07]. Appendix D describes this technique in detail.

3. This same technique may be used to regularize the noisy measurements of $E(\mathbf{q})$ in a very convenient way (see Appendix D).

As an additional advantage, it has been proved in [Des07] that SH are also eigenfunctions for the FRT operator, so this representation may be combined with the FRT to obtain a non–parametric estimator not depending on any model for $E(\mathbf{q})$. Therefore, two of the three requirements introduced in Section 6.1 are simultaneously met. In precise terms:

$$\begin{aligned} &\mathscr{G}\left\{\widetilde{Y}_{2l}^m(\mathbf{g})\right\}(\mathbf{r}) = 2\pi(-1)^l\frac{(2l-1)!!}{(2l)!!}\widetilde{Y}_{2l}^m(\mathbf{r}) \\ \Rightarrow\ &\mathscr{G}\left\{\mathscr{H}(\mathbf{g})\right\}(\mathbf{r}) = \sum_{l=0}^{\infty}\sum_{m=-2l}^{2l} 2\pi(-1)^l\frac{(2l-1)!!}{(2l)!!}C_{2l}^m\,\widetilde{Y}_{2l}^m(\mathbf{r}), \end{aligned} \tag{6.17}$$

where the standard notation for the double factorial has been used (see eq. (D.5) in Appendix D). Hence, the FRT of $\mathscr{H}$ is approximated correcting the coefficients of its truncated SH expansion.

6.5 The Orientation Probability Density Transform

6.5.1 Estimation of the OPDF from the attenuation signal

The aim of this Chapter is to find an estimator fulfilling the requisites mentioned in Section 6.1. To avoid (at least in part) the need for any unrealistic assumption on the behavior of $E(\mathbf{q})$ outside $\mathscr{S}$, i.e. for values other than $E(q_0\mathbf{g})$, it was stated in Section 6.3 the convenience of using the FRT. Therefore, the estimator to be designed will be written in the form:

$$\begin{aligned}\hat{p}(\mathbf{r}) &= \mathscr{G}\{\mathscr{H}(\mathbf{g})\} = \iint_{\mathscr{S}'} \delta(\mathbf{r}^T\mathbf{g})\mathscr{H}(q_0\mathbf{g})d\mathbf{g}\\ &= \frac{2\pi}{q_0}\int_0^\infty\int_0^{2\pi} J_0(2\pi q_0 u)\left(\int_{-\infty}^{\infty}\widetilde{\mathscr{P}}(u,\beta,\vartheta)d\vartheta\right)ud\beta du\\ &\simeq 2Z'(q_0)\int_0^\infty \mathscr{P}(R\mathbf{r})dR, \end{aligned} \tag{6.18}$$

for a given function $\mathscr{H}$, whose Fourier transform is $\mathscr{P}$, to be determined next.

The second important requisite is that the estimator $\hat{p}(\mathbf{r})$ has to represent true probabilistic information. From Section 6.2 it remains evident that $p(\mathbf{r})$ is necessarily an estimator of the OPDF. Attending its definition in eq. (6.5), the last equality in eq. (6.18) imposes the condition:

$$\mathscr{P}(\mathbf{R}) = R^2 P(\mathbf{R}). \tag{6.19}$$

Since $\mathscr{P}$ and $\mathscr{H}$ form a Fourier transform pair, eq. (6.18) may be used to infer the condition $\mathscr{H}$ has to meet. From basic Fourier theory, it is clear that this condition is:

$$\mathscr{H}(\mathbf{q}) = \frac{-1}{4\pi^2}\left(\frac{\partial^2}{\partial {q_1}^2}+\frac{\partial^2}{\partial {q_2}^2}+\frac{\partial^2}{\partial {q_3}^2}\right)E(\mathbf{q}), \tag{6.20}$$

where $\mathbf{q} = [q_1, q_2, q_3]^T$. The right–hand side of eq. (6.20) is clearly proportional to the Laplacian of E, so joining the previous results, it yields:

$$\hat{p}(\mathbf{r}) = \Phi(\mathbf{r}) \simeq \frac{1}{2Z'(q_0)}\mathscr{G}\left\{\frac{-1}{4\pi^2}\Delta E(\mathbf{q})\right\}(\mathbf{r}) = C\cdot\mathscr{G}\{\Delta E(\mathbf{q})\}(\mathbf{r}), \tag{6.21}$$

for a negative constant C. As stated in Section 6.4, the radial symmetry of the problem advices to represent the Laplacian operator in spherical coordinates:

$$\begin{aligned}\Delta E &= \frac{1}{q^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial E}{\partial\theta}\right)+\frac{1}{q^2\sin^2\theta}\frac{\partial^2 E}{\partial\phi^2}+\frac{1}{q^2}\frac{\partial}{\partial q}\left(q^2\frac{\partial E}{\partial q}\right)\\ &= \frac{1}{q^2}\Delta_b E+\frac{1}{q^2}\Delta_q E. \end{aligned} \tag{6.22}$$

The problem reduces to estimate the Laplacian of the attenuation signal $E(\mathbf{q})$ in the sampled sphere $\mathscr{S}$ to compute its FRT. The operator may be decomposed in a term proportional to the Laplace–Beltrami operator and a radial–dependent term. The following Sections are respectively centered in the estimation of each of them. It results very illustrative in this analysis to consider the closed–form expressions resulting for Gaussian diffusion propagators. They are surveyed in Appendix E.

6.5.2 Estimation of Δ_b

The third condition imposed in Section 6.1 was to get rid of any parametric model in order to represent both the diffusion signal and the diffusion propagator. It will be satisfied by expressing both functions with the SH formulation given in eq. (6.16). For the diffusion signal:

$$E(q_0\mathbf{g}) = \sum_{l=0}^{\infty} \sum_{m=-2l}^{2l} C_{2l}^m Y_{2l}^m(\mathbf{g}), \tag{6.23}$$

where the coefficients C_{2l}^m are theoretically obtained as in eq. (6.16). In practice, the technique presented in Appendix D is used instead for a truncated version of eq. (6.23) up to order $L = 2\hat{L}$. The fundamental property of SH of being eigenfunctions of the Laplace–Beltrami operator may be easily exploited to calculate the first term in eq. (6.22):

$$\frac{1}{q^2}\Delta_b E(q_0\mathbf{g}) = \frac{1}{q^2}\sum_{l=0}^{\infty} \sum_{m=-2l}^{2l} 2l(2l+1) C_{2l}^m Y_{2l}^m(\mathbf{g}). \tag{6.24}$$

Note that only the values of $E(\mathbf{q})$ inside $\mathscr{S}$ are required to evaluate eq. (6.24).

6.5.3 Estimation of Δ_q

The estimation of the radial term in eq. (6.22) is more involved. Only the information on the behavior of $E(\mathbf{q})$ inside $\mathscr{S}$ is available, but the computation of Δ_q requires the estimation of radial derivatives. To circumvent this problem, making a **weak** assumption on the behavior of $E(\mathbf{q})$ in the neighborhood of q_0 is mandatory:

$$\begin{aligned}\frac{\partial E(\mathbf{q})}{\partial q} \equiv \frac{\partial E(q,\mathbf{g})}{\partial q} &= \frac{\partial}{\partial q}\exp\left(-4\pi^2\tau q^2 D(\mathbf{q})\right) = \frac{\partial}{\partial q}\exp\left(-4\pi^2\tau q^2 D(q,\mathbf{g})\right)\\ &= -4\pi^2\tau q\left(2D(q,\mathbf{g}) + q\frac{\partial}{\partial q}D(q,\mathbf{g})\right)\exp\left(-4\pi^2\tau q^2 D(q,\mathbf{g})\right)\\ &\simeq -8\pi^2\tau q D(q,\mathbf{g})\exp\left(-4\pi^2\tau q^2 D(q,\mathbf{g})\right).\end{aligned} \tag{6.25}$$

It is required to assume that $\partial D/\partial q \ll 2D/q$ for $q = q_0$; this implies to suppose a slow variation of the ADC compared to the exponential decay of the diffusion signal $E(\mathbf{q})$. Compared to the strong assumption of a constant ADC for all q (like for the DOT or higher order tensors), this one is much weaker for two reasons:

1. It does not require the ADC to be constant, but only to vary much slower than the diffusion signal itself.
2. This condition has to hold only in a **local sense** in an environment of q_0, and not for all q.

Nonetheless, the need for a (weak) assumption is a clear drawback of the estimator proposed compared to Q–Balls. It is important to note that Q–Balls is the only estimator ever reported not needing any supposition at all. Besides, it will be shown in Section 6.6.5 that in fact eq. (6.25) represents a very realistic behavior, so the premise of a local sense–slowly varying ADC is rarely an issue. Another important point is that the weakness of the assumption is such thanks to the use of the FRT.

Finally, the same reasoning may be used for the second order derivatives to compute Δ_q:

$$\begin{aligned}\Delta_q E(q_0\mathbf{g}) &\simeq -8\pi^2\tau q_0^2 D(q_0,\mathbf{g})\left(3 - 8\pi^2\tau q_0^2 D(q_0,\mathbf{g})\right)E(q_0\mathbf{g})\\ &= 2\log E(q_0\mathbf{g})\left(3 + 2\log E(q_0\mathbf{g})\right)E(q_0\mathbf{g}).\end{aligned} \tag{6.26}$$

It is interesting to notice the formal similarity between this approximation and the exact closed–form expression for the case of Gaussian diffusion propagators,

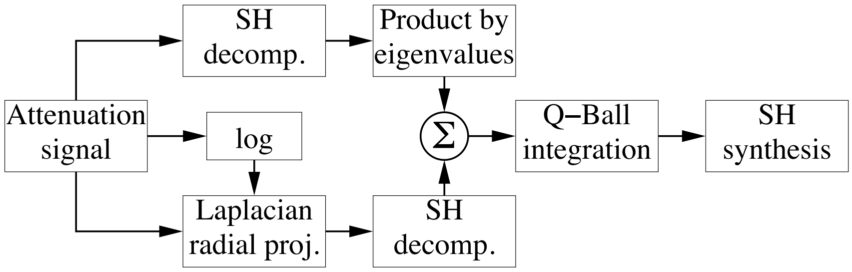


Figure 6.4: Summary of the proposed OPDF estimator, the so–called OPDT. Details on the computation of each block may be found in Appendix D.

given by eq. (E.4) in Appendix E. In fact, eq. (6.26) is exact for the tensor model, if the ADC $D(q_0\mathbf{g})$ is substituted by the quadratic form $\mathbf{g}^T\mathscr{D}\mathbf{g}$, which evidences the consistency of the estimator of Δ_q.

6.5.4 Computation of the Orientation Probability Density Transform

Once the Laplacian of $E(\mathbf{q})$ has been calculated, the estimator in eq. (6.21) may be computed as summarized in Fig. 6.4. The diffusion signal is expressed in the SH basis, so that the computation of Δ_b is trivial. Eq. (6.26) is used to compute Δ_q, and the result is expressed as a SH sum. The linearity of eq. (6.16) is exploited to compute the SH coefficients of the Laplacian of $E(\mathbf{q})$, and eq. (6.17) is used to compute its FRT. All these processes may be performed as matrix operations, as described in Appendix D, to yield the closed–form expression in eq. (D.17). The implementation given by this equation for the estimator in eq. (6.21) defines the so–called Orientation Probability Density Transform (OPDT) [TV09d].

The numerical implementation of the OPDT is based on the Q–Balls implementation in [Des07], so the OPDT shares a number of characteristics with this other approach, mainly its robustness and fast computation. Under a pure numerical point of view, the OPDT may be seen as a kind of contrast–enhanced Q–Balls: it is computed as the FRT of the diffusion signal corrected with the Laplacian operator, which has the effect of sharpening the orientation function computed by direct FRT, i.e. the ODF. The OPDT has the following **advantages**:

1. It computes true probabilistic information, since it is designed to estimate the OPDF.
2. It only needs a very weak assumption on the behavior of $E(\mathbf{q})$.
3. It does not rely on any parametric model, since its numerical implementation is based on SH.
4. Its computation is fast and robust, due to the regularized matrix operations described in Appendix D.
5. It is able to yield accurate results even for commonly used low values of q_0, as it is shown in Section 6.6.

Note that the three first points correspond to the requirements imposed in Section 6.1. On the other hand, some **drawbacks** are present in this novel estimator:

1. Although weak, an assumption on $E(\mathbf{q})$ outside $\mathscr{S}$ is still required.
2. The OPDT emphasizes fast varying SH basis functions, so it is more prone to noise artifacts than regular Q–Balls.

3. The OPDT is based on the FRT, so it shares the same problem of Q–Balls regarding the blurring of the orientation information (see eq. (6.10)).

The first problem, as shown in Section 6.6.5, is not very critical since the premise used is quite realistic. As a consequence, the third condition in Section 6.1 is mostly met. The second problem may be worked around using the unbiased filtering techniques proposed in Chapter 5, so it is not so critical either. The last point is more important. Besides the aforementioned blurring of the estimated OPDF, an additional difficulty arises from the use of the FRT, which is described in Section 6.5.5, and which motivates the study carried out in Chapter 7 to search alternatives to the FRT.

6.5.5 Issues in the computation of the FRT

The blurring due to the Bessel kernel in the FRT (see Section 6.3) has an additional side effect in the OPDT due to J_0 is not positive (see Fig. 6.3, right). Consider the expression of the Laplacian for the tensor model given by eq. (E.3) in Appendix E. Due to the positive–definite character of $\mathscr{D}$, it may be possible to find that, for the directions of maximum diffusion (largest $\|\mathscr{D}\mathbf{q}\|$):

$$-8\pi^2\tau\left(-8\pi^2\tau\|\mathscr{D}\mathbf{q}\|^2+\mathrm{trace}(\mathscr{D})\right)E(\mathbf{q})>0, \tag{6.27}$$

and so certain directions of $\mathbf{q}$, once the factor $-1/(4\pi^2)$ is applied, contribute to the FRT integral with negative values. This means that the integrals corresponding to certain directions $\mathbf{r}$ of the OPDF orthogonal to the directions $\mathbf{q}$ of maximum diffusion may become negative. Therefore, the modulus of the OPDT estimation has to be computed in all cases. The orientations orthogonal to those of maximum diffusion correspond to directions where no fiber bundles are present, so the distortion in the OPDF is not so important. However, it is clear that an estimator of a PDF should not yield negative values, so this artifact induced by the FRT is the most important drawback of the OPDT. Yet, for q_0 large enough, eq. (6.27) might become negative for all possible orientations, so the OPDT would be useless. For Q–Balls, the FRT averages the values of $E(\mathbf{q})$, which are always positive, along an equator, so Q–Balls yield always positive estimates. Hence, this artifact is a consequence of combining the estimation of the OPDF (instead of the ODF) with the Bessel kernel–pondered integral inherent to the FRT.

6.6 Numerical simulations

This Section is devoted to the quantitative evaluation of the OPDT. In particular, the capability of resolving fiber crossings, and correctly determining their directions in a variety of scenarios, is assessed. This methodology has become a common benchmark to test the accuracy in the resolution of complex fiber architectures in HARDI [Des07, Öza06, Prč08, Tou07, Tou08, Tuc04]. To simulate, as far as it is possible, a realistic environment, there is a number of typical scanning parameters of interest:

- The value of b_0 (related to q_0 as: $b_0=4\pi^2\tau q_0^2$, see eq. (2.31)).
- The number of sampling gradient directions in $\mathscr{S}$, N.
- The noise power σ_n^2, see Chapter 3.

6.6.1 Methodology

Generation of the synthetic DWI signal

It is assumed that the voxel under study contains a known number of differentiated fiber bundles in well–defined directions. Besides, the time between diffusion pulses Δ is short enough so that no water exchange occurs between different fibers. Under these conditions, the probability of diffusion may be seen as the mixture of F independent Gaussian diffusion propagators [Des07, Tou07, Tuc04]:

$$E(q_0\mathbf{g}) = \sum_{f=1}^{F} p_f \exp\left(-b_0 \mathbf{g}^T \mathscr{D}_f \mathbf{g}\right), \tag{6.28}$$

where p_f are the partial volume fractions associated to each fiber direction and $\mathscr{D}_f$ their respective covariance matrices (diffusion tensors). In practice, the value $F = 2$ is used in all experiments except for Section 6.6.4. Despite this is a well–accepted methodology, in [Öza06, Prč08] the diffusion signal is generated with a model based on isotropic diffusion inside a bounded cylinder. This is a simplified model for the microscopic diffusion inside a nervous cell, but not for the macroscopic behavior measured by the scanner (see the introduction in Section 2.5 for a detailed discussion on this topic). This is the reason why eq. (6.28) is used here instead. Although it has not been previously reported in the literature, note that using eq. (6.28) for validation could bias the results in favor of those methods assuming a model based on Gaussian mixtures (spherical deconvolution, multi–tensor models and continuous mixtures of Gaussians, for example). It is important to realize that eq. (6.28) is a highly over–simplified model of the complex architectures present in the white matter.

Noise corruption

Although the effect of noise may be palliated using the techniques presented in Chapter 5, studying the impact of MRI noise in the estimation of fiber directions is a good complement to the analysis carried out in Chapter 4. To that end, some noisy scenarios have been simulated corrupting the ideal DWI signal given by eq. (6.28) with Rician noise[2]. In a real situation, $E(q_0\mathbf{g})$ is computed as the quotient between the diffusion–weighted gradient images and the baseline image, which are both Rician–distributed. The statistics of noise in this case have not been studied. For this reason, the same assumption as in Chapter 4 is taken, i.e. the baseline is supposed to be noise–free, so $E(q_0\mathbf{g})$ will be Rician–distributed itself. The noisy signal will be generated as:

$$\widetilde{E}(q_0\mathbf{g}) = \sqrt{\left(E(q_0\mathbf{g}) + n_c\right)^2 + n_s^2}, \tag{6.29}$$

for n_c, n_s two independent Gaussian processes with zero mean and variance σ_n^2. The PSNR (see Chapter 5) is used to represent the noise power in each experiment.

Comparison to other related estimators

The OPDT does not rely on any parametric model, so the comparison with spherical deconvolution, multi–tensor models, or continuous mixtures of Gaussians is

---

[2] Although it was stated in Chapter 3 that some modern protocols for pMRI may produce non–central Chi distributed signals, the Rician distribution is currently the most often found case, so the study is centered on it.

discarded. Higher order tensors model the ADC and not the probability of water diffusion, so they are not a good target to compare to either. Persistent angular structures make use of a very unrealistic assumption, which is a characteristic the OPDT tries to avoid. There are two remaining methods:

- The DOT [Öza06] shares a number of characteristics with the OPDT: its numerical scheme is based on SH, and it is very fast and robust to compute as well. The main difference between them is that the DOT requires to assume a constant ADC for all q, so the comparison will be very useful to stress the importance of relaxing this premise (see Section 6.5.3). Besides, note that the DOT is an estimator for $\Upsilon(\mathbf{r})$ and not $\phi(\mathbf{r})$.
- Q–Balls is the closest approach to the OPDT, mainly due to the use of the FRT, common to both of them. Q–Balls does not require any kind of assumption, but on the other hand it is an estimator for $\Psi(\mathbf{r})$, which does not represent actual probabilities. Like for the DOT, comparing this characteristic with the OPDT, which estimate true probabilities, will be very useful to stress the importance of this feature of the OPDT. The implementation of Q–Balls is that of [Des07].

Both estimators are non–parametric (due to the use of SH expansions), and may be computed in a very fast and robust manner. As a final remark, neither the OPDT, nor the other two estimators to be evaluated, assume a diffusion propagator writable in terms of a mixture of Gaussians, so the use of eq. (6.28) does not introduce a bias in the comparison.

Simulation parameters

The tensors $\mathscr{D}_f$ in eq. (6.28) are generated with eigenvalues $[1.8, 0.2, 0.2] \cdot 10^{-3}\text{mm}^2/s$, which is a typical value for HARDI evaluation [Bas96]. The principal eigenvectors are in the XY plane: for the first tensor, the principal eigenvector forms an angle of $(90-\alpha)/2$ degrees with the X axis, and for the second one an angle of $(90-\alpha)/2$ degrees with the Y axis. The directions of both fiber–bundles form an angle of α degrees, with partial volume fractions $p_f = 1/2$ each. $E(q_0\mathbf{g})$ is sampled for N antipodal symmetric directions inside $\mathscr{S}$. Five different scenarios are considered: **S-1**, with $b_0 = 1200\text{s/mm}^2$ and $N = 50$; **S-2**, with $b_0 = 1200\text{s/mm}^2$ and $N = 100$; **S-3**, with $b_0 = 3000\text{s/mm}^2$ and $N = 50$; **S-4**, with $b_0 = 3000\text{s/mm}^2$ and $N = 100$; **S-5**, with $b_0 = 3000\text{s/mm}^2$ and $N = 200$. The PSNR for noisy scenarios is indicated where needed.

Parameters used for the estimators

SH expansions up to order $L = 6$ ($H = 28$ SH coefficients) have been used in all cases. For the regularization parameter (see Appendix D), $\lambda = 0.006$ is a good trade–off between smoothness and representation capability [Des07]. For the DOT, the only parameter (apart from L) which has to be fixed is the radius R_0 for which $\Upsilon(\mathbf{r}) = P(R_0\mathbf{r})$ is computed. According to [Öza06, Prč08], a higher value ($R_0 = 25\mu\text{m}$) is preferable for noise–free, ideal scenarios. In a real situation, a more conservative compromise of $R_0 = 12\mu\text{m}$ has to be considered, so this latter value has been used in all cases to achieve a fair comparison.

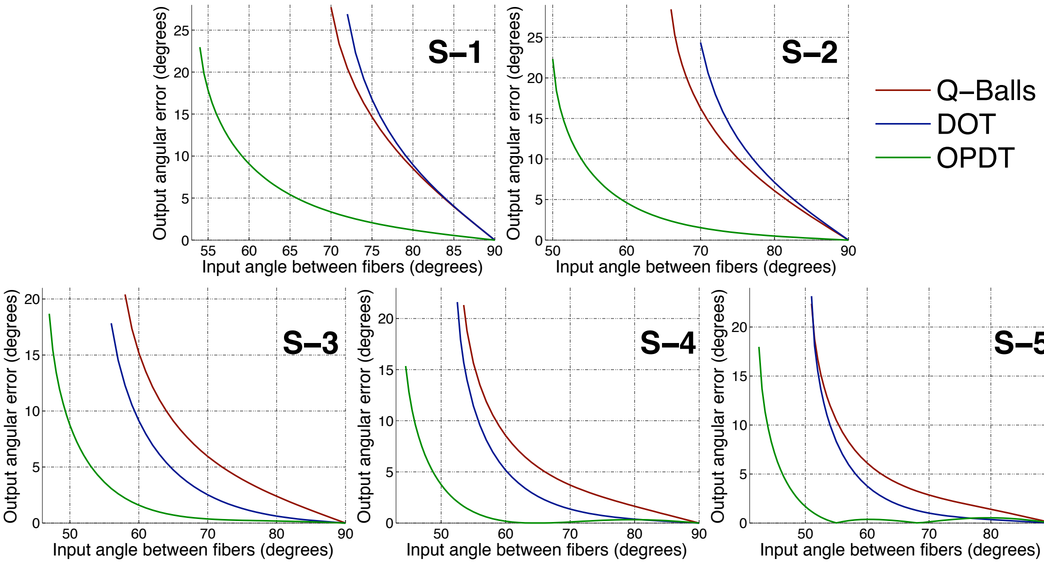


Figure 6.5: Mean error in the angle of detected fibers as a function of the original angle between fibers. Fiber directions are identified as local maxima (computed form SH coefficients up to the numerical precision of the machine) of the orientation function in each case. Curves stop when the estimators are not able to detect the fiber crossings (there are not two maxima in the orientation function).

6.6.2 Capability of resolving two crossing fibers

Fig. 6.5 represents the average angular error between the ground–truth directions and the directions given by each estimator. Obviously, the closer the angle between the ground–truth fiber bundles, the more difficult to distinguish between them and the higher the error. Fibers are detected as local maxima of the orientation function. In the case of the OPDT, this function ($\Phi(\mathbf{r})$) is a true PDF, so this methodology represents a maximum likelihood estimator for fiber directions[3]. This property does not hold for Q–Balls or the DOT, since they do not estimate actual probabilities.

The OPDT is able to resolve fibers crossing in closer directions in all scenarios, and the difference is more important for lower b_0 values. The OPDT is more accurate than Q–Balls in all cases. The DOT is more accurate than the OPDT only for higher values of b_0 and fibers crossing in large angles, and yet its advantage is quite subtle. Comparing Q–Balls and the DOT, their performance is quite similar for low values of b_0, but the DOT performs better for higher b_0 (especially with few gradients); similar results have been reported in [Prč08]. Note that all estimators are able to better resolve fiber crossings in closer directions for higher b_0, since a greater b_0 augments the contrast in the diffusion profile. On the contrary, using more gradient directions improves the results only marginally: the minimum detectable angle is practically the same, and the improvement on the angular error is negligible. As a final comment, note the anomalous behavior of the OPDT for certain crossing angles (55^o to 75^o degrees depending on the number of gradients) with b_0=3000mm/s^2: the angular error decreases as the fiber bundles get closer. This phenomenon is hardly visible with few gradient directions (scenario **S–3**), but it is perfectly visible in scenarios **S–4** and **S–5**. Yet, for **S–5**, two minima are visible in the corresponding curve. This artifact will be thoroughly discussed in Section 6.6.5.

As an additional result, Fig. 6.6 shows glyph representations of the orientation

---

[3] Using the maximum of the orientation information function is a standard methodology to infer fiber directions [Blo08]. However, note that it cannot be fully justified unless true probabilities are estimated, as it is the case for the OPDT.

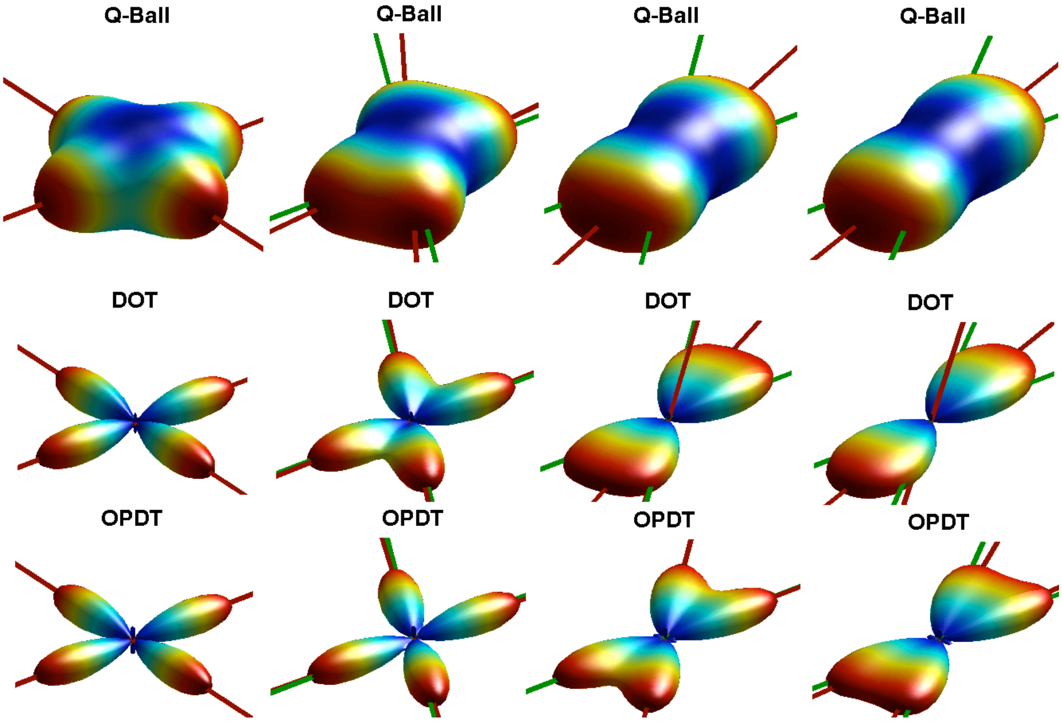


Figure 6.6: Glyph representations of the estimated orientation functions for **S–4** with Q–Balls (top), the DOT (middle), and the OPDT (bottom), for crossing angles, from left to right: 90^o, 60^o, 45^o and 40^o. Green lines correspond to the ground–truth, while red lines correspond to the estimated directions.

functions for the three estimators (scenario **S–4**). The OPDT yields more accurate estimates and is able to detect the two fibers when the other estimators show only one maximum (Q–Balls for 45^o and 40^o) or a maximum which does not correspond to any of the true fiber populations (DOT for 45^o and 40^o). The OPDF estimated by the OPDT is sharper than the information inferred with the other techniques.

From this Section it may be concluded that the OPDT is preferable: **first** to resolve fibers crossing in low angles, and **second** for low values of b_0; the **third** conclusion is that the OPDF is in general able to yield more accurate results and a higher angular contrast, with the additional advantage of being a true PDF.

6.6.3 Behavior in the presence of noise

The robustness of the estimators against noise is tested in the experiments depicted in Fig. 6.7. For the scenarios with lower b_0 values, PSNR=13.3 is used, and for the higher b_0 values PSNR=5.0: since it is easier to detect fibers for higher b_0, the estimators are able to tolerate greater noise powers as b_0 increases. Anyway, note that the PSNR used represent severe noisy scenarios. Even so, all the estimators are still able to satisfactorily infer fiber directions, though obviously noise worsens their accuracy. As opposed to the noise–free scenarios, in this case using more gradient directions notably improves both the minimum angle detectable and the accuracy. Note that increasing b_0 improves the result even when the SNR dramatically decreases.

The advantage of using the OPDT is especially noticeable for lower b_0. For higher b_0, the OPDT is also more accurate and show a better resolution capability than the DOT, except for vary large crossing angles. Contrary to the behavior shown in Fig. 6.5, Q–Balls is now more accurate for higher b–values and large

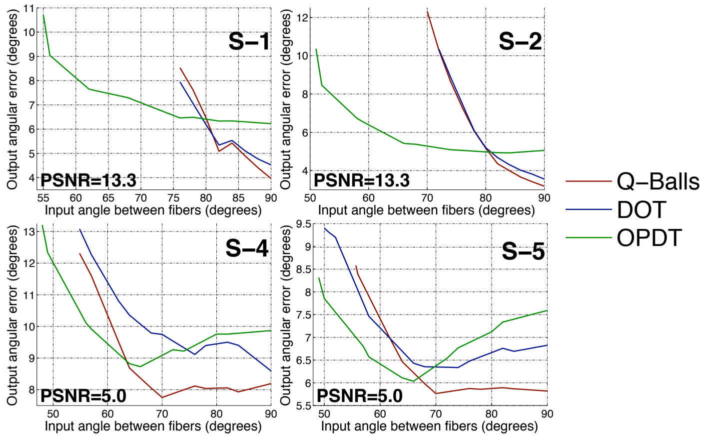


Figure 6.7: Mean error in the angle of the detected fibers as a function of the original angle between fibers in the presence of noise. Results are the average of 100 Montecarlo trials. A failure is considered (and thus the representation of the curves stop) whenever the estimator is not able to detect the two fibers in the 50% of the trials.

input angles. Qualitatively speaking, Q–Balls works by averaging the attenuation signal in a whole equator perpendicular to the direction of interest; therefore, this estimator should be more robust to noise than, for example, the OPDT, which is based on derivatives computed on the attenuation signal, accentuating the effect of noise. Nevertheless, the signal averaging inherent to Q–Balls produces an angular blurring of the orientation information, which justifies the fact that the OPDT is able to recover lower crossing angles even in noisy scenarios. For low b–values, this blurring is more important than the effect of noise, and both the OPDT and the DOT perform better than Q–Balls for all crossing angles.

Note that a value of $b_0 = 1200\text{s}/\text{mm}^2$ is more realistic in practical applications than $3000\text{s}/\text{mm}^2$, so the advantage of the OPDT remains clear. Yet, for $b_0 = 3000\text{s}/\text{mm}^2$, the OPDT is more accurate than the DOT, except for very large crossing angles, and has a better resolution capability. Compared to Q–Balls, the OPDT is able to resolve crossing angles far lower. An analogous representation to that of Fig. 6.6 is shown in Fig. 6.8 for scenario **S–4**, with PSNR=5. Consistently with the results in Fig. 6.7, Q–Balls gives closer estimates to the true fibers, but its representation of the fiber populations is less intuitive and more blurred, failing to find the two populations before the DOT or the OPDT do, so the final remark is the better representation capability of the OPDT and the DOT when compared to Q–Balls.

6.6.4 Resolution of more complex architectures

For illustrative purposes, Fig. 6.9 shows a situation with three crossing fiber bundles; the directions are: $(\theta_1,\phi_1)=(\pi/2,0)$, $(\theta_2,\phi_2)=(\pi/2,\pi/2)$ and $(\theta_3,\phi_3)=(3\pi/20,\pi/2)$; the eigenvalues of $\mathscr{D}_i$ are: $[2,0.2,0.3]\cdot 10^{-3}$, $[1.8,0.4,0.3]\cdot 10^{-3}$ and $[2,0.1,0.1]\cdot 10^{-3}$ mm^2/s. The scenarios tested are **S–2** with PSNR=13.3 and **S–5** with PSNR=5. Q–Balls completely blurs the orientation information, and it is not able to distinguish the three fibers. The DOT fails to recover the three fibers for **S–2**, but for **S–5** it is able to detect all of them with a mean error of 18^o. The OPDT is able to yield acceptable results in both situations, with respective mean errors of 10^o and 16^o. This result evidences that the OPDT is more accurate and robust as well for more complex architectures. As mentioned in Sections 6.6.2 and 6.6.3, its advantage is clear for lower b_0 values.

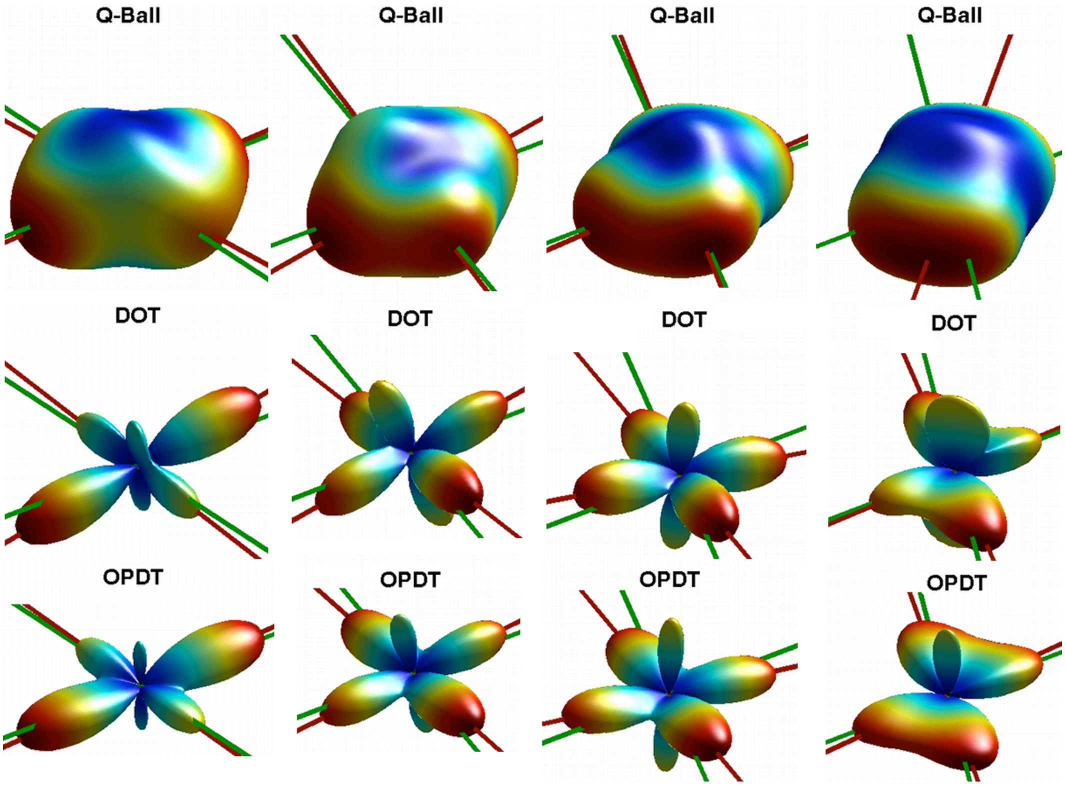


Figure 6.8: Glyph representations of the estimated orientation functions in a noisy scenario, for **S–4** with Q–Balls (top), the DOT (middle), and the OPDT (bottom), for crossing angles, from left to right: 90^o, 75^o, 65^o and 60^o. Rician noise with PSNR=5 has been added to $E(\mathbf{q})$.

6.6.5 Accuracy in the approximation of the Laplacian

It is shown in Fig. 6.10 a pseudo–color representation of $(2D/q)/(\partial D/\partial q)$ (see eq. (6.25)) and Δ_b/Δ_q (see eq. (6.22)), computed as described in Appendix E for 2 and 3 crossing fibers. The figure shows averaged values of these quotients for all b_0 between 1000 and 3000 s/mm^2. In the most of the (θ,ϕ) space, $(2D/q)/(\partial D/\partial q)$ is greater than 10, and yet Δ_b represents more than the 90% of the total value of Δ, so the error is negligible. The greatest errors (small $(2D/q)/(\partial D/\partial q)$) are commit-

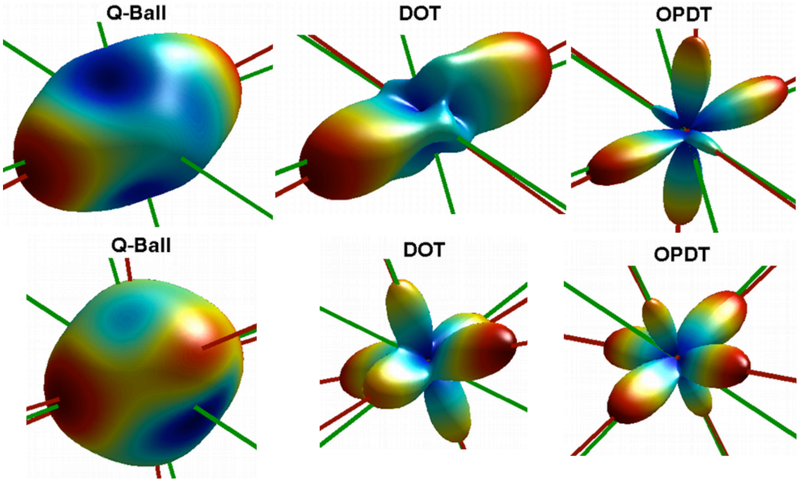


Figure 6.9: Glyph representations of the estimated orientation functions with three crossing fibers, for **S–2** with PSNR=13.3 (top), and **S–5** with PSNR=5 (bottom).

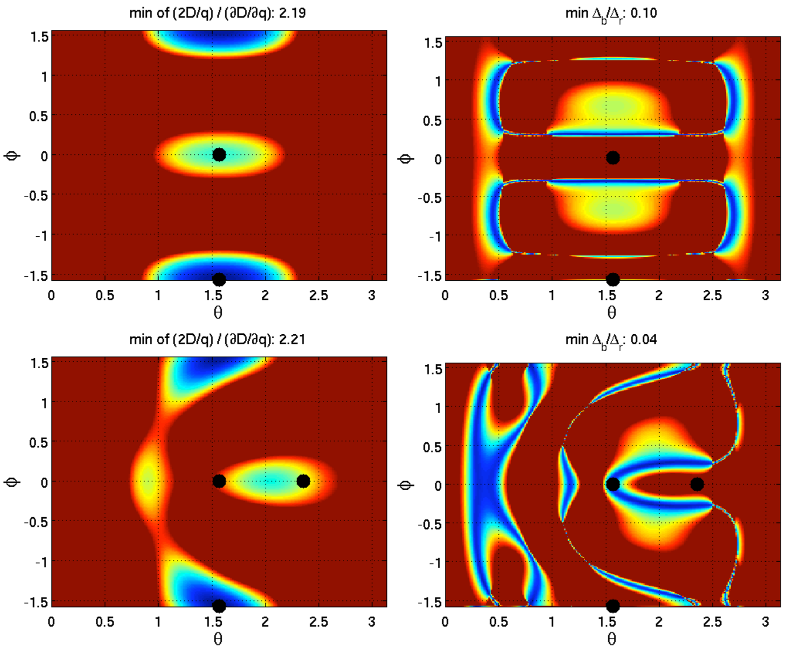


Figure 6.10: Pseudo–color representation of $(2D/q)/(\partial D/\partial q)$ (error in the estimation of Δ_q, left) and Δ_b/Δ_q (relative importance of Δ_b, right) for 2 (top) and 3 (bottom) crossing fibers. Half the (θ,ϕ) space is represented since the attenuation signal is symmetric. Fiber directions are represented as black spots, and values greater than 10 have been clipped to this value in all cases.

ted in the directions of the fiber bundles. However, with 2 fibers these directions correspond to zones where Δ_b is clearly greater, so the overall error is small. With 3 fibers, there are directions where $(2D/q)/(\partial D/\partial q)$ is small, and Δ_b/Δ_q is less than 10, but the minima of $(2D/q)/(\partial D/\partial q)$ never overlap with minima of Δ_b/Δ_q. It follows, first, that more complicated fiber architectures yield greater errors, and, second, that the relative error due to Δ_q remains moderated compared to the total value of $\Delta = q_0^2(\Delta_q + \Delta_b)$.

To support these conclusions, Fig. 6.11 shows the relative error in the estimation of the Laplacian as a function of the angle between 2 crossing fibers. The error decreases in all cases as the fibers get closer, since the structure of $E(\mathbf{q})$ is simpler. Compared to Fig. 6.5, note that, with the OPDT, the angular error anomaly decreases for fibers in angles in a range from 55^o to 75^o, which may be explained by the fact that for these angles the relative error in Δ_q dramatically decreases. This artifact, however, depends on the number of sensitizing gradients, and is not very critical. Yet, when the fibers get too close, the estimator is not able to resolve them, and the error increases not depending on the error in Δ_q, so that

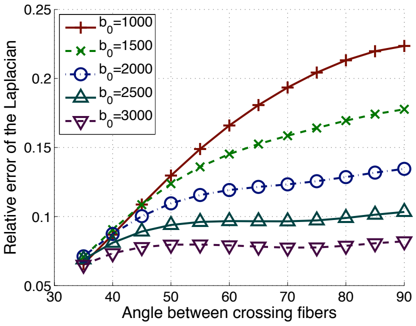


Figure 6.11: Relative error (averaged for $N = 100$ gradient directions) in the estimation of the Laplacian as a function of the angle between 2 crossing fibers.

the artifact is completely removed.

Note that the error in Δ_q is more important with lower b_0. The assumption explained in Section 6.5.3 implies that the exponential decay of $E(\mathbf{q})$ is much faster than the change in the shape of the ADC. For lower values of b_0, the shape of $D(\mathbf{q})$ changes faster, but for higher b_0, $E(\mathbf{q})$ follows an asymptotic mono–exponential decay, so the ADC varies very slowly compared to this exponential decay. Nonetheless, it has been previously shown that the OPDT is more advantageous for lower b_0, so it may be concluded that the error in the estimation of Δ_q is negligible.

6.7 Conclusion

The OPDT is a simple, robust, but yet very accurate estimator of the probability density of fiber populations in the white matter. Its complexity is comparable to that of Q–Balls or the DOT, but it can better resolve fiber crossings and represent fiber populations. Besides, it has a correct probabilistic interpretation: Q–Balls only estimates radial projections of true probability densities, and the DOT estimates only the value of the probability density at a given radius R_0. On the contrary, the OPDT is an estimator of true marginal probability densities (OPDF), which turns out to be an important property. In effect, it allows to interpret the localization of maxima in the OPDF as a maximum likelihood estimate of fiber populations. The technique proposed is more advantageous for low b_0 values, so its computation is not very restrictive on the machinery used to acquire the diffusion images. However, as any other estimator ever reported in the literature, its capability to resolve very close crossing fibers improves with higher b_0; for this kind of images, the SNR is highly degraded, so the adverse effect of noise may overwhelm the higher contrast in some cases. Nevertheless, the OPDT has shown to be equally very robust to noise.

Compared to other methods not directly evaluated in this Chapter, the OPDT is the only one fulfilling all the requisites stated in Section 6.1 for an estimator of fiber populations: contrary to higher order tensors, it estimates true probabilities of diffusion, and does not require to assume a constant ADC for all q. Regarding multi–tensor models, spherical deconvolution or continuous mixtures of Gaussians, the OPDT does not require to suppose any particular model for $E(\mathbf{q})$, which may take any form. In particular, unlike spherical deconvolution, the OPDT does not require the prior estimation of a deconvolution kernel, whose accuracy clearly impacts the ability of the estimator to infer fiber directions. In this sense, the approach in [Des09] is formally very similar to the novel ODPT, since it is based on the sharpening of the Q–Balls–computed ODF by the weighting of its SH parameters. But this method requires to estimate a deconvolution kernel, which makes it more similar to all other spherical deconvolution approaches [And05,Tou07,Tou08] than to the OPDT. Finally, compared to Persistent angular structures, the OPDT does not require any unrealistic assumption, and it is able to estimate true probabilities.

The OPDT has a number of drawbacks as well:

- It still requires to make a (weak) assumption on the behavior of $E(\mathbf{q})$ outside $\mathscr{S}$. As shown in Section 6.6.5, this is rarely an issue, so it may be considered that in fact the OPDT accomplish the three requirements of interest.

- The weakness of this premise is achieved with the use of the FRT. Besides the well–known blurring in the computed OPDF due to the use of this technique, an additional drawback was detailed in Section 6.5.5, which basically consists in the lack of a positive–definite character of the estimated orientation functions.

This latter issue is without any doubt the main problem of the OPDT, and therefore the following Chapter is centered on finding a solution for this problem.

7

Integral approximations to probabilistic fiber orientation estimators: beyond the Funk–Radon Transform

In the previous Chapter, a novel technique for the inference of fiber orientations, the OPDT, was introduced. It was shown that the key for the calculation of probability densities is the integration of the weighted diffusion propagator along the direction of interest. This integration was translated to an integral in the dual Fourier domain (i.e. the $\mathbf{q}$–space) by means of the FRT. With Q–Balls estimation, the FRT allows to approximate radial integrals without making any assumption on the behavior of the attenuation signal outside the sampled sphere $\mathscr{S}$. With the novel OPDT, it allows to highly relax such assumptions, so that it is only necessary to suppose that the shape of the ADC is approximately constant in a local sense. The drawback of FRT estimation is the blurring of the radial integrals, which are estimated as integrals inside Bessel–shaped tubes. Moreover, with the OPDT there is an additional disadvantage: the orientation information may become negative under certain conditions. The present Chapter is devoted to avoid this limitation. First, it is shown that, although it was empirically shown that the negativity problem with the OPDT is not so critical in practical situations, the calculation based on the FRT is an important theoretical limitation for this estimator. Then, an alternative solution based on Stokes' theorem is given: the assumption of the model is the same as with the OPDT, i.e. a slow varying ADC, but the blurring due to the FRT is heavily reduced. This technique may be applied both to Q–Balls and the OPDT. Moreover, the relatively high accuracy of the DOT encourages to use the premise in this model (a constant ADC for all q) to obtain alternative expressions. As a result, four novel estimators, two for $\Psi(\mathbf{r})$ and two for $\Phi(\mathbf{r})$, are introduced. They represent the main original contributions in this Chapter, and are summarized in Table 7.1. All of them admit a representation in terms of the FRT of a given function obtained from the attenuation signal. This way, the FRT is no longer an estimator itself, but it becomes a very powerful mathematical tool for HARDI techniques without the artifacts inherent to its direct use for the calculation of radial integrals.

7.1 Introduction

In the previous Chapter, the FRT has been proved a highly valuable tool for HARDI estimators of fiber populations. It allows the characterization of diffusion propagators from the values of the attenuation signal sampled only at a sphere of radius q_0. For this reason, it has been successfully used in the literature in Q–Balls estimators [Tuc03, Tuc04] and other recent techniques [Des09]. Despite of its FRT grounding, this latter approach is conceptually a spherical deconvolution method, hence basically different from Q–Balls. Therefore, it makes sense to consider these two approaches as different techniques, and the precedent for a part of the work developed in this thesis.

Concerning the OPDT presented in Chapter 6, the FRT is also the central technique for the inference of orientation information. Besides, the calculation of the underlying line integral becomes simpler with the SH expansions–based implementation proposed in [Des07]. All these considerations evidence the high interest of this tool for HARDI studies.

It was previously stated that the main virtue of the FRT is the allowance of any assumption in the behavior of the attenuation signal outside the sampled sphere (or, at least, any *strong* assumption, for the case of the OPDT). Due in part to this feature, the OPDT outperforms other estimators, like the DOT, in many situations. Nonetheless, the FRT also shows a number of drawbacks which are shared by all the estimators based on this technique. As argued in the previous Chapter, there are two main limitations:

1. The FRT is used to estimate the radial integral along a direction $\mathbf{r}$ by the integration along an equator perpendicular to this direction. This estimation is not exact, introducing an important blurring described by Bessel functions–shaped kernels (see eq. (6.10)).

2. For the case of the OPDT, the negativity of the estimated orientation information under certain conditions is a problematic side effect of the estimation of the OPDF instead of the ODF.

While the former is not very critical (both Q–Balls and the OPDT has been shown to yield accurate estimates regardless on this source of error, and yet the spherical deconvolution approach in [Des09] has been proved very useful), the latter may become an important limitation to the use of the OPDT. In particular, consider eq. (E.3), which represents the Laplacian of the attenuation signal if a simple tensor model is considered. This expression is reproduced here for convenience:

$$\Delta E(\mathbf{q}) = -8\pi^2\tau\left(-8\pi^2\tau\|\mathscr{D}\mathbf{q}\|^2 + \mathrm{tr}(\mathscr{D})\right)E(\mathbf{q}). \tag{7.1}$$

As discussed in the previous Chapter, the positive definite character of $\mathscr{D}$ ensures that, for $\|\mathbf{q}\|$ large enough, the previous equation may yield positive values. Since the OPDT is computed as the FRT of the Laplacian multiplied by a negative constant, negative values of the OPDF may appear. The work–around used with the OPDT is to eliminate these negative values by simply changing them by 0. This problem becomes more serious if the b–value (and, consequently, q_0) is increased, since the second term inside the brackets in eq. (7.1) (the trace of the diffusion tensor) does not depend on b, but the first one increases linearly with b (or quadratically with q_0). Yet, if the b–value is arbitrarily increased, eq. (7.1) may become positive for all $\mathbf{q}$, making the OPDT useless. Although the experimental work carried out in Chapter 6 has shown that for practical b–values and common tensor configurations this is not an issue, this problem is an important theoretical limitation of the OPDT.

The aim of this Chapter is to give an adequate solution to this problem. Such a solution must not compromise the fulfillment of the requisites defined in Section 6.1 for fiber orientation estimators. It is desirable that the proposed solution estimates actual probabilities with the weakest possible assumptions on the attenuation signal, and that it is still non–parametric. The statement of this problem drives to the circulation–based OPDT derived in Section 7.4. However, it is shown in Sections 7.3 and 7.5 that relaxing these conditions in some way drives to estimators which may be useful for some tasks.

7.2 Line integrals as flux integrals in the q–space

The aim in this Chapter is to compute the line integral in eq. (6.5) (or the one in eq. (6.6)), avoiding the blurring introduced by the FRT, by means of an alternative representation in the transformed Fourier domain. To that end, the methodology is as follows:

1. Deduce the exact expression relating the integral of the diffusion propagator along a given direction $\mathbf{r}$ (possibly weighted by the Jacobian of the spherical coordinates system) to its equivalent in the Fourier domain. It translates to an integral in the orthogonal plane to $\mathbf{r}$, Π (see Fig. 7.1).
2. Approximate this calculation by the integral in a bounded domain, i.e., inside the disk Ω delimited in Π by the sphere $\mathscr{S}$.
3. Relate the latter integral to a line integral along the equator of $\mathscr{S}$ described by its intersection with the plane Π (i.e., along the boundary Γ of Ω).

In what follows, each of these steps is motivated and discussed.

7.2.1 Representation of line integrals in the Fourier domain

Consider the auxiliary cylindrical coordinates system described in Fig. 6.2, where the 'z' axis has been aligned with the direction of interest, $\mathbf{r}$. The integral of the diffusion propagator $P(\mathbf{R})$ (or, alternatively, the weighted diffusion propagator $R^2P(\mathbf{R})$) may be easily computed from eq. (6.6) as:

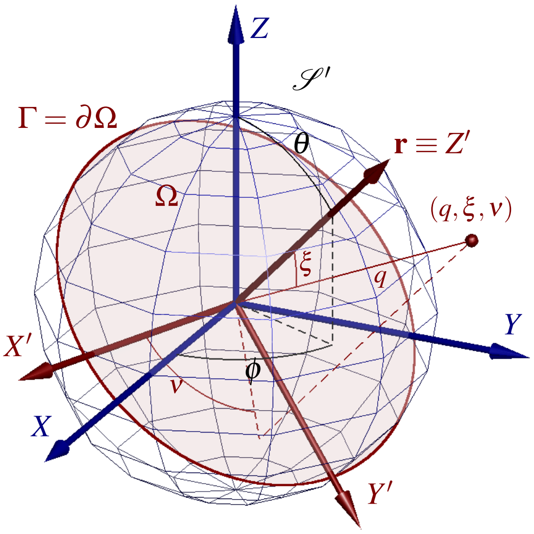
$$\begin{aligned}
\Psi(\mathbf{r}) &= \frac{1}{2}\int_{-\infty}^{\infty} P(R\mathbf{r})dR = \frac{1}{2}\int_{-\infty}^{\infty} \widetilde{P}(u,\beta,\vartheta)d\vartheta\bigg|_{u=0}\\
&= \frac{1}{2}\int_{-\infty}^{\infty}\left(\int_{0}^{\infty}\int_{0}^{2\pi}\int_{-\infty}^{\infty}\widetilde{E}(t,\alpha,z)\exp\left(-j2\pi\left(t\cdot 0\cdot\cos\left(\alpha-\beta\right)+z\vartheta\right)\right)\; t\; dz\; d\alpha\; dt\right)d\vartheta\\
&= \frac{1}{2}\int_{0}^{\infty}\int_{0}^{2\pi}\int_{-\infty}^{\infty}\widetilde{E}(t,\alpha,z)\left(\int_{-\infty}^{\infty}\exp\left(-j2\pi z\vartheta\right)d\vartheta\right)\; t\; dz\; d\alpha\; dt\\
&= \frac{1}{2}\int_{0}^{\infty}\int_{0}^{2\pi}\left(\int_{-\infty}^{\infty}\widetilde{E}(t,\alpha,z)\delta\left(z\right)dz\right)\; t\; d\alpha\; dt = \frac{1}{2}\int_{0}^{\infty}\int_{0}^{2\pi}\widetilde{E}(t,\alpha,0)\; t\; d\alpha\; dt, \qquad (7.2)
\end{aligned}$$

where $\widetilde{P}$ and $\widetilde{E}$ are respectively the cylindrical coordinates representations of P and E, and δ stands for Dirac's delta distribution. The last term in eq. (7.2) is clearly the integral of the attenuation signal in the 'xy' plane of the auxiliary system (note the presence of the Jacobian of the transformation, t, due to the polar representation of the integral). Translated to real–world coordinates, this plane Π is the plane orthogonal to $\mathbf{r}$ containing the origin. This same reasoning may be applied to the weighted diffusion propagator, so that the following equalities hold:

$$\Psi(\mathbf{r}) = \frac{1}{2}\iint_{\Pi_{\perp\mathbf{r}}} E(\mathbf{q})d\Pi. \qquad (7.3)$$

$$\Phi(\mathbf{r}) = \frac{-1}{8\pi^2}\iint_{\Pi_{\perp\mathbf{r}}} \Delta E(\mathbf{q})d\Pi. \qquad (7.4)$$

In the previous equations, the notation $\Pi_{\perp\mathbf{r}}$ has been used to make it explicit that the plane Π is orthogonal to $\mathbf{r}$. Comparing eq. (7.2) with the FRT approximation in eq. (6.10), the former represents the exact calculation of the radial integral, while the latter is only an approximation. With the FRT, the integral in the whole plane Π is substituted by the integral in a circumference contained on it. Obviously, this



Parametrization of Γ:

$$\Gamma \equiv \left\{ \begin{array}{c} q = q_0 \\ \xi = \pi/2 \\ \nu = \nu \end{array} \right. , \ \nu \in [0, 2\pi)$$

Correspondences between the spherical coordinates in the **q**–space and the dual **R**–space:

$$\begin{array}{c} q \leftrightarrow R \\ \xi \leftrightarrow \zeta \\ \nu \leftrightarrow \upsilon \end{array}$$

Figure 7.1: Auxiliary **spherical** coordinates system for the computation of the integrals inside $\Omega \subset \Pi$. The equator $\Gamma = \partial\Omega$ is orthogonal to the direction of interest, $\mathbf{r}$. The dual **R**–space with respect to the Fourier transform is represented by the triplet (R, ζ, υ).

is a very rough approximation which introduces large errors in the estimation, as shown in Chapter 6. On the contrary, eq. (7.4) may be used to compute the exact ODF, without any error. In particular, the main problem of the OPDT, i.e. its negativity under some circumstances, is avoided.

Unfortunately, conventional HARDI techniques do not allow the characterization of $E(\mathbf{q})$ (or other related functions) in the whole plane Π, but only in its intersection with the sphere $\mathscr{S}$. The work–around with the FRT is to integrate the corresponding function only along the equator Γ in Fig. 7.1. Instead, an alternative approach is derived in what follows.

7.2.2 Approximation of line integrals as integrals in a disk

The integrals in eqs. (7.3) and (7.4) cannot be explicitly computed only from values sampled at $\mathscr{S}$ over the corresponding equator. Instead, the integral inside the disk $\Omega \subset \Pi$, whose boundary $\Gamma = \partial\Omega$ is the intersection of $\mathscr{S}$ with Π, may be considered. A new auxiliary coordinates system has to be introduced at this point, which has been depicted in Fig. 7.1. In this case, the problem better fits a spherical system. Like in Fig. 6.2, the coordinates system is rotated so that the direction of interest, $\mathbf{r}$, corresponds to the colatitude angle $\xi = 0$. Therefore, the disk Ω is easily represented as: $\Omega = \{(q,\xi,\nu) | 0 \leq q \leq q_0, \xi = \pi/2, 0 \leq \nu < 2\pi\}$, and:

$$\Psi(\mathbf{r}) \simeq \frac{1}{2}\iint_{\Omega} E(\mathbf{q}) d\Omega = \frac{1}{2}\int_0^{q_0}\int_0^{2\pi} \tilde{E}(q, \frac{\pi}{2}, \nu) q d\nu dq. \qquad (7.5)$$

$$\Phi(\mathbf{r}) \simeq \frac{-1}{8\pi^2}\iint_{\Omega} \Delta E(\mathbf{q}) d\Omega = \frac{-1}{8\pi^2}\int_0^{q_0}\int_0^{2\pi} \widetilde{\Delta E}(q, \frac{\pi}{2}, \nu) q d\nu dq. \qquad (7.6)$$

In general, the decay of the attenuation signal follows an exponential law, so the most of the energy of $E(\mathbf{q})$ (or $\Delta E(\mathbf{q})$) is concentrated inside Ω. For this assumption to hold, it is not necessary to assume a mono–exponential decay of $E(\mathbf{q})$: it has been shown that the attenuation signal is accurately described by a multi–exponential model [Nie96, Öza06, Shi09]. Moreover, the physics of diffusion impose a fast decay of $E(\mathbf{q})$ as q_0 increases (see Chapter 2), so it may be assumed that the error committed neglecting the integral outside Ω is small enough compared to the integral inside this subset.

Of course, the error in the computation decreases as q_0 (or, alternatively, the b–value) increases, since the subset of Π for which the integral is neglected is smaller, comprising more attenuated tails of $E(\mathbf{q})$. Compared to the computation based on the FRT, note that the problem of negative estimations is noticeable especially for higher b–values, so the approach proposed in this Chapter is especially attractive for OPDT–like estimators. Nevertheless, there is a side problem in using the integral inside Ω: an exponential model (multi–exponential, in fact) ensures that only a small fraction of the actual value of the integral is neglected; but the exponential law, and consequently the error, is different for each direction $\mathbf{g}$. In case this error is not negligible (for example, if those directions of minimum diffusion are considered), it may introduce a certain distortion in the estimated orientation function.

Apart from these considerations, it remains to compute the integrals inside Ω, and for this task only sampled values along Γ (those available in the HARDI data set) may be used. The next Section is centered on this topic.

7.2.3 Flux integrals in the Fourier domain and Stokes' theorem

Stokes' theorem naturally arises when relating the integral inside a bounded surface Ω to the integral along its boundary Γ. It states that the circulation of a vector field, $\mathbf{F}$, along a closed curve Γ in $\mathbb{R}^3$ equals the flux integral of its curl, $\mathbf{H}$, across the surface Ω enclosed by Γ. For the particular situation depicted in Fig. 7.1, Stokes' theorem reads:

$$\int_{\Gamma} \mathbf{F}\cdot\mathbf{t}\, d\Gamma = \iint_{\Omega} (\nabla\times\mathbf{F})\cdot\mathbf{n}\, d\Omega = \iint_{\Omega} \mathbf{H}\cdot\mathbf{n}\, d\Omega, \tag{7.7}$$

where $\mathbf{t}$ is a unit vector tangent to Γ at each point and $\mathbf{n}$ is a unit vector normal to Ω; the orientations of $\mathbf{t}$ and $\mathbf{n}$ are related through the *right hand rule*. $\nabla\times\mathbf{F}$ is the conventional notation for the curl of a vector field. Using the auxiliary spherical coordinates system in Fig. 7.1, the domains of integration, the tangent and normal unit vectors, and the Jacobian determinants for the integration may be explicitly computed:

$$\int_{\Gamma} \mathbf{F}\cdot\mathbf{t}\, d\Gamma = \int_0^{2\pi} \widetilde{\mathbf{F}}(q_0,\frac{\pi}{2},\nu)\cdot\mathbf{e}_{\nu}\; q_0 d\nu = \int_0^{2\pi} \widetilde{F_{\nu}}(q_0,\frac{\pi}{2},\nu) q_0 d\nu; \tag{7.8}$$

$$\iint_{\Omega} \mathbf{H}\cdot\mathbf{n}\, d\Omega = \int_0^{q_0}\int_0^{2\pi} \widetilde{\mathbf{H}}(q,\frac{\pi}{2},\nu)\cdot\left(-\mathbf{e}_{\xi}\right) q d\nu dq = -\int_0^{q_0}\int_0^{2\pi} \widetilde{H_{\xi}}(q,\frac{\pi}{2},\nu) q d\nu dq, \tag{7.9}$$

where $\mathbf{e}_{\nu}$ and $\mathbf{e}_{\xi}$ are the unit vectors orthogonal to the isosurfaces of ν and ξ, respectively; $\widetilde{\mathbf{F}}$ and $\widetilde{\mathbf{H}}$ are the representations of $\mathbf{F}$ and $\mathbf{H}$ in this system, and F_{ν} and H_{ξ} their corresponding projections on $\mathbf{e}_{\nu}$ and $\mathbf{e}_{\xi}$. The normal vector $\mathbf{n}$ is aligned with $-\mathbf{e}_{\xi}$ obeying the *right hand rule*.

Identifying corresponding terms in eqs. (7.5) (respectively, eq. (7.6)), and (7.9), it is clear that the attenuation signal E (or respectively, its Laplacian) has to be identified with the component of $\mathbf{H}$ in the direction of $\mathbf{e}_{\xi}$. Since the other components of $\mathbf{H}$ (H_q and H_{ν}) are orthogonal to $\mathbf{e}_{\xi}$, they do not contribute to the flux integral in eq. (7.7); hence, the estimators in eqs. (7.5) and (7.6) may be written as the flux across Ω of the following vector fields:

$$\widetilde{\mathbf{H}^{\Psi}} = \widetilde{H_{\xi}^{\Psi}}\; \mathbf{e}_{\xi} = -\frac{1}{2}\widetilde{E}(q,\xi,\nu)\; \mathbf{e}_{\xi}. \tag{7.10}$$

$$\widetilde{\mathbf{H}^{\Phi}} = \widetilde{H_{\xi}^{\Phi}}\; \mathbf{e}_{\xi} = \frac{1}{8\pi^2}\widetilde{\Delta E}(q,\xi,\nu)\; \mathbf{e}_{\xi}. \tag{7.11}$$

For each direction of interest $\mathbf{r}$, the auxiliary system, and thus $\mathbf{e}_\xi$, changes. As a consequence, $\mathbf{H}$ is different for each $\mathbf{r}$. Fortunately, this problem can be overcome in a very simple way, as stated in Section 7.6. For a particular direction $\mathbf{r}$, $\mathbf{H}$ has to be expressed in terms of the curl of a given field $\mathbf{F}$, to be computed using the expression of this differential operator in spherical coordinates. Since $\mathbf{H}$ follows the direction of $\mathbf{e}_\xi$ at each point, only this component has to be considered, yielding:

$$\nabla\times\widetilde{\mathbf{F}}=\widetilde{\mathbf{H}}\Rightarrow\frac{1}{q\sin(\xi)}\left(\frac{\partial\widetilde{F_q}}{\partial\nu}-\sin\xi\frac{\partial q\widetilde{F_\nu}}{\partial q}\right)=\widetilde{H_\xi}. \tag{7.12}$$

The previous expression is only the particularization of the curl operator to the component in $\mathbf{e}_\xi$. Besides, the boundary Γ of Ω is tangent at each point to the unit vector $\mathbf{e}_\nu$ (see Fig. 7.1). From eq. (7.8), it is evident that the components of $\mathbf{F}$ orthogonal to $\mathbf{e}_\nu$ (i.e., F_q and F_ξ) are perpendicular to the trajectory of integration, and thus they do not contribute to the circulation along Γ. Consequently, the term F_q in eq. (7.12) may be dropped down. Under a different point of view, note that F_q should be periodic in ν to ensure the differentiability of $\mathbf{F}$ (otherwise, a singularity in the $X'Z'$) plane of Fig. 7.1 would appear). As a result, its derivative would be periodic in ν as well, and the corresponding integral in a whole period would be 0. Then, from eq. (7.9) it becomes evident that the contribution of this term to the flux across Ω would simply be 0. Thinking of any of these interpretations, it is clear that eq. (7.12) is equivalent to:

$$\widetilde{H_\xi}(q,\xi,\nu)=-\frac{1}{q}\frac{\partial q\widetilde{F_\nu}(q,\xi,\nu)}{\partial q}. \tag{7.13}$$

For each H_ξ in eqs. (7.10) or (7.11), eq. (7.13) has to be solved for F_ν. The estimators of the corresponding orientation information functions are computed using eq. (7.8):

$$\Psi(\mathbf{r})\simeq\frac{1}{2}\iint_{\Omega}E(\mathbf{q})d\Omega=\int_{\Gamma}\mathbf{F}^{\Psi}\cdot\mathbf{t}\,d\Gamma=\int_0^{2\pi}\widetilde{F_\nu^{\Psi}}(q_0,\frac{\pi}{2},\nu)q_0d\nu, \tag{7.14}$$

where $\widetilde{F_\nu^{\Psi}}$ is the result of solving eq. (7.13) for $\widetilde{F_\nu}$, with $\widetilde{H_\xi}$ corresponding to $\widetilde{\mathbf{H}^{\Psi}}$ in eq. (7.10). Accordingly, for $\Phi(\mathbf{r})$:

$$\Phi(\mathbf{r})\simeq\int_0^{2\pi}\widetilde{F_\nu^{\Phi}}(q_0,\frac{\pi}{2},\nu)q_0d\nu. \tag{7.15}$$

Both estimators can now be written in terms of integrals along Γ, where Π may be characterized from the HARDI data set. To this point, the only source of error in the estimators is the part of the integral over Π which is neglected (the subset outside the disk Ω), since the line integrals in eqs. (7.14) and (7.15) are identical to their bi–dimensional counterparts in eqs. (7.5) and (7.6). It only remains to solve eq. (7.13) for F_ν to explicitly evaluate the estimators. The following Sections give particular solutions to this problem for known estimators.

7.3 Circulation–based Q–Balls

Q–Balls is an estimator of the ODF $\Psi(\mathbf{r})$, i.e the integral of the diffusion propagator along $\mathbf{r}$. It may be generalized with the technique presented in this Chapter particularizing eq. (7.13) for $\widetilde{\mathbf{H}^{\Psi}}$ in eq. (7.10):

$$-\frac{1}{2}\widetilde{E}(q,\xi,\nu)=-\frac{1}{q}\frac{\partial q\widetilde{F_\nu^{\Psi}}(q,\xi,\nu)}{\partial q}\Rightarrow\frac{q}{2}\widetilde{E}(q,\xi,\nu)=\frac{\partial q\widetilde{F_\nu^{\Psi}}(q,\xi,\nu)}{\partial q}. \tag{7.16}$$

Like for the OPDT presented in Chapter 6, a radial derivative appears in the statement of this problem. The solution given in Section 6.5.3 was to assume that the ADC was nearly constant for q_0 in a local sense. The experiments carried out in Section 6.6.5 suggested that in fact this is a reasonable assumption. With this premise, the previous equation may be explicitly solved:

$$\begin{aligned}
\widetilde{F_\nu^\Psi}(q,\xi,\nu) &= \frac{1}{q}\int_0^q \frac{u}{2}\widetilde{E}(u,\xi,\nu)du = \frac{1}{q}\int_0^q \frac{u}{2}\exp\left(-4\pi^2\tau u^2\widetilde{D}(u,\xi,\nu)\right)du \\
&\simeq \frac{1}{q}\int_0^q \frac{u}{2}\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)du = \frac{1}{2q}\left[\frac{\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)}{-8\pi^2\tau\widetilde{D}(q_0,\xi,\nu)}\right]_0^q \\
&= \frac{1-\exp\left(-4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu)\right)}{16\pi^2\tau q\widetilde{D}(q_0,\xi,\nu)}. \qquad (7.17)
\end{aligned}$$

The integration of the partial derivatives equation has been performed between 0 and q because Stokes' theorem needs certain regularity conditions on $\mathbf{F}$ to hold. In particular, a singularity at $q=0$ is not allowed. Since q appears in the denominator, it is necessary to ensure that the numerator tends to 0 for $q=0$ to avoid a pole in the function. In other words, the integration between 0 and q is equivalent to choose an integration constant such that for any value of ξ and ν the numerator of eq. (7.17) tends to 0.

It is important to stress that this **is not** equivalent to assume that the ADC is constant between 0 and q_0. Eq. (7.17) represents an approximation to the primitive of the function to integrate only for $q=q_0$. Amongst the infinite number of primitives of such a function, the one whose value at $q=0$ is 0 is chosen. Although the final result is numerically equivalent to suppose that the ADC is constant (and equal to the actual ADC for $q=q_0$) between 0 and any given q, there is a deep conceptual difference: if the ADC is assumed constant for all q, the final result in eq. (7.17) holds for every q (not only q_0). On the contrary, if the ADC is only assumed to vary slowly near q_0, eq. (7.17) is only valid for q_0. Since the integration of F_ν to compute the estimator requires to know its value only for q_0, the latter assumption is enough. This is a main conceptual difference with other HARDI estimators like the DOT or higher order tensors.

With this result, eq. (7.14) can now be evaluated to compute the final estimator:

$$\begin{aligned}
\Psi(\mathbf{r}) &\simeq \int_0^{2\pi}\frac{1-\exp\left(-4\pi^2\tau {q_0}^2\widetilde{D}(q_0,\frac{\pi}{2},\nu)\right)}{16\pi^2\tau q_0\widetilde{D}(q_0,\frac{\pi}{2},\nu)}q_0 d\nu \\
&= \frac{q_0^2}{4}\int_0^{2\pi}\frac{1-\widetilde{E}(q_0,\frac{\pi}{2},\nu)}{-\log\left(\widetilde{E}(q_0,\frac{\pi}{2},\nu)\right)}d\nu, \qquad (7.18)
\end{aligned}$$

which is an alternative to Q–Balls partially avoiding the blurring due to the FRT. This source of error is palliated at the expense of introducing a weak assumption on the behavior of the attenuation signal outside the sampled sphere (as stated in Chapter 6, one of the main advantages of Q–Balls is the riddance of any assumption). Besides, the computation of the circulation along Γ is only an approximation of the actual ODF, since it is equivalent to the integral inside the disk $\Omega\subset\Pi$. However, this approximation is better than simply integrating E along Γ, which is the FRT approach.

7.4 Circulation–based OPDT

The OPDT is based on the integration of the Laplacian of the attenuation signal. Integrating ΔE in the plane Π orthogonal to the direction of interest $\mathbf{r}$, the true marginal probability of diffusion along this direction (the OPDF) is computed. In spherical coordinates, the Laplacian may be represented as described in Section 6.5.1:

$$\Delta E(\mathbf{q}) = \frac{1}{q^2}\Delta_q E(\mathbf{q}) + \frac{1}{q^2}\Delta_b E(\mathbf{q}) = \frac{1}{q^2}\frac{\partial}{\partial q}\left(q^2\frac{\partial E(\mathbf{q})}{\partial q}\right) + \frac{1}{q^2}\Delta_b E(\mathbf{q}). \qquad (7.19)$$

The first term in eq. (7.19) is the radial part, whose representation in the auxiliary system of Fig. 7.1 is exactly the same, since the radial coordinate q is the same in both systems. Besides, the integral of the radial term over the plane Π may be explicitly computed integrating by parts:

$$\begin{aligned}
\iint_\Pi \frac{1}{q^2}\widetilde{\Delta_q E}(q,\xi,\nu)d\Pi &= \int_0^{2\pi}\int_0^\infty \frac{1}{q^2}\frac{\partial}{\partial q}\left(q^2\frac{\partial \widetilde{E}(q,\frac{\pi}{2},\nu)}{\partial q}\right) q\,dq\,d\nu \\
&= \int_0^{2\pi}\left(\left[q\frac{\partial \widetilde{E}(q,\frac{\pi}{2},\nu)}{\partial q}\right]_0^\infty + \int_0^\infty \frac{\partial \widetilde{E}(q,\frac{\pi}{2},\nu)}{\partial q}dq\right)d\nu \\
&= \int_0^{2\pi}\left(\lim_{q\to\infty} q\frac{\partial \widetilde{E}(q,\frac{\pi}{2},\nu)}{\partial q} + \lim_{q\to\infty}\widetilde{E}(q,\frac{\pi}{2},\nu) - \widetilde{E}(0,\frac{\pi}{2},\nu)\right)d\nu \\
&= \int_0^{2\pi}(0+0-1)\,d\nu = -2\pi, \qquad (7.20)
\end{aligned}$$

where the limits for $q \to \infty$ and $q = 0$ are computed taking into account only the physics of diffusion; for $q = 0$, no diffusion gradient is applied, and therefore there is no attenuation. For $q \to \infty$, the same discussion of Section 7.2.2 regarding the exponential behavior of $E(\mathbf{q})$ holds, so eq. (7.20) has general validity without any simplification.

The integral related to the Laplace–Beltrami operator cannot be computed in a closed form, so the same methodology as in Section 7.3 has to be used. In this case, eq. (7.13) particularizes to the following expression:

$$\frac{1}{8\pi^2 q^2}\widetilde{\Delta_b E}(q,\xi,\nu) = -\frac{1}{q}\frac{\partial q\widetilde{F_\nu^\Phi}(q,\xi,\nu)}{\partial q} \Rightarrow \frac{-1}{8\pi^2 q}\widetilde{\Delta_b E}(q,\xi,\nu) = \frac{\partial q\widetilde{F_\nu^\Phi}(q,\xi,\nu)}{\partial q}. \qquad (7.21)$$

To solve the previous equation for $\widetilde{F_\nu^\Phi}$, it is necessary to assume once again a slow variation of the ADC, due to the existence of radial derivatives. Compared to the OPDT presented in Chapter 6, now it is the term depending on the angular part of the Laplacian (the Laplace–Beltrami operator) which requires to make extra assumptions on the attenuation signal. On the other hand, the term depending on the radial Laplacian may be exactly computed without any error (yet, it is not necessary to restrict the integral to the disk Ω, so absolutely no error is committed), as opposed to the case of the OPDT. Nonetheless, the assumption required for this approach is exactly the same as for the OPDT, so the new technique does not introduce new sources of error *per se*.

The same methodology used to solve eq. (7.16) may be reproduced with eq. (7.21):

$$\begin{aligned}
\widetilde{F_\nu^\Phi}(q,\xi,\nu) &= \frac{-1}{8\pi^2 q}\int_0^q \frac{1}{u}\widetilde{\Delta_b E}(u,\xi,\nu)du = \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b \exp\left(-4\pi^2\tau u^2\widetilde{D}(u,\xi,\nu)\right)}{u}du \\
&\simeq \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b \exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)}{u}du, \qquad (7.22)
\end{aligned}$$

where the same discussion about the integration between 0 and q of the previous Section holds now. However, the computation of this integral is not trivial, since the division by u introduces a pole which is not integrable for $u = 0$. It is shown in Appendix F that in fact the integral in eq. (7.22) is always convergent due to the computation of the Laplace–Beltrami operator, which assures that a zero of a greater order than the pole is present. Indeed, it is shown that eq. (7.22) may be written in the form:

$$\begin{aligned}\widetilde{F^{\Phi}_{\nu}}(q,\xi,\nu) &\simeq \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b \exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)}{u}du\\ &= \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b\left(\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)-1\right)}{u}du\\ &= \frac{-1}{8\pi^2 q}\Delta_b\int_0^q \frac{\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)-1}{u}du, \end{aligned}\tag{7.23}$$

where the order of the Laplace–Beltrami operator and the radial integral can be exchanged since the new integral is convergent. Although the integral in eq. (7.23) is not trivial, it may be written in terms of the exponential integral E_i [Abr72], whose main properties are reviewed in Appendix F:

$$\begin{aligned}\widetilde{F^{\Phi}_{\nu}}(q,\xi,\nu) &\simeq \frac{-1}{8\pi^2 q}\Delta_b\int_0^q \frac{\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)-1}{u}du\\ &= \frac{-1}{8\pi^2 q}\Delta_b\left[\frac{1}{2}\mathrm{E}_i\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)-\log(u)\right)\right]_0^q\\ &= \frac{-1}{16\pi^2 q}\Delta_b\mathrm{E}_{in}(4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu)), \end{aligned}\tag{7.24}$$

where E_{in} is the non–singular exponential integral, whose definition, together with some asymptotic properties, are given in Appendix F. Finally, eq. (7.24) may be casted into eq. (7.15) to compute the estimate of the integral of the angular Laplacian. This result, together with eq. (7.20), allows to compute the final estimator of the OPDF:

$$\begin{aligned}\Phi(\mathbf{r}) &\simeq \int_0^{2\pi}\frac{-1}{16\pi^2 q_0}\Delta_b\mathrm{E}_{in}\left(4\pi^2\tau q_0^2\widetilde{D}(q_0,\frac{\pi}{2},\nu)\right)q_0 d\nu + \frac{-1}{8\pi^2}(-2\pi)\\ &= \frac{-1}{16\pi^2}\int_0^{2\pi}\Delta_b\mathrm{E}_{in}\left(4\pi^2\tau q_0^2\widetilde{D}(q_0,\frac{\pi}{2},\nu)\right)d\nu + \frac{1}{4\pi}\\ &= \frac{-1}{16\pi^2}\int_0^{2\pi}\Delta_b\mathrm{E}_{in}\left(-\log\left(\widetilde{E}(q_0,\frac{\pi}{2},\nu)\right)\right)d\nu + \frac{1}{4\pi}. \end{aligned}\tag{7.25}$$

This estimator is an alternative to the OPDT intended to reduce the blurring due to the computation of the FRT. It has two main sources of error: first, the angular part of the Laplacian is not integrated in the whole plane Π, but only over the subset $\Omega \subset \Pi$. This distortion may be reduced increasing the b–value (i.e. increasing the radius q_0 of the sampling sphere). Compared to the OPDT, this is a clear advantage, since the problems of the OPDT are especially noticeable for larger b–values. Second, the ADC has to be supposed constant, in a local sense, in an environment of q_0. Since this is the same assumption required to compute the OPDT, the new approach does not introduce a new source of error in this sense. Besides, eq. (7.25) is a true estimator for the OPDF without any kind of normalization: suppose that $E(\mathbf{q})$ is constant for all $\mathbf{q}$; in this case the computation of the Laplace–Beltrami operator yields simply 0, so the estimator reduces to a constant value of $1/4\pi$, i.e. to a uniform variable in the extent of the sphere of radius 1. Therefore, the scaling of the OPDT by a factor q_0^2 is avoided.

7.5 Alternative estimators of orientation information

The representation of line integrals as flux integrals, together with the application of Stokes' theorem, have an additional advantage. When solving eqs. (7.16) and (7.21), it was stated that the weak assumption of a slow varying ADC ensures the validity of the obtained functions only for $q = q_0$; but if the ADC is assumed to be constant for all q, the solutions given by eqs. (7.17) and (7.24) remain valid for all q. These results may thus be used to compute the limit of the flux integrals when q approaches infinity, which has to yield the integral in the whole plain Π. This way, the error due to the reduction of the integration domain is completely removed at the expense of imposing a stronger condition on the behavior of $E(\mathbf{q})$. In other words, the main weakness of the FRT (the blurring in the radial integral) is completely avoided, but at the same time its main advantage (the absence of any restriction on $E(\mathbf{q})$) is lost. This premise has been successfully applied in other related estimators such as the DOT or higher order tensors, so it should render estimations which are at least adequate.

For the estimator proposed in Section 7.3, and from eq. (7.17), the circulation $C(\widetilde{F_\nu^\Psi},\Gamma)$ for a given value of q may be calculated as:

$$\begin{aligned} C(\widetilde{F_\nu^\Psi},\Gamma) &\simeq \int_0^{2\pi} \frac{1-\exp\left(-4\pi^2\tau q^2\widetilde{D}(q_0,\frac{\pi}{2},\nu)\right)}{16\pi^2\tau q\widetilde{D}(q_0,\frac{\pi}{2},\nu)} q d\nu \\ &= \int_0^{2\pi} \frac{1-\exp\left(-4\pi^2\tau q^2\widetilde{D}(q_0,\frac{\pi}{2},\nu)\right)}{16\pi^2\tau\widetilde{D}(q_0,\frac{\pi}{2},\nu)} d\nu. \end{aligned} \tag{7.26}$$

This expression holds approximately for all q if the ADC is assumed approximately constant for all q. Therefore, it make sense to take its limit when q tends to infinity to compute the integral in the whole plain Π[1]:

$$\Psi(\mathbf{r}) = \lim_{q\to\infty} C(\widetilde{F_\nu^\Psi},\Gamma) \simeq \int_0^{2\pi} \frac{d\nu}{16\pi^2\tau\widetilde{D}(q_0,\frac{\pi}{2},\nu)} = \frac{q_0^2}{4}\int_0^{2\pi} \frac{-d\nu}{\log\left(\widetilde{E}(q_0,\frac{\pi}{2},\nu)\right)}. \tag{7.27}$$

The same reasoning may be done for the circulation–based OPDT, and from the second equality in eq. (7.25):

$$\begin{aligned} \Phi(\mathbf{r}) &= \lim_{q\to\infty} C(\widetilde{F_\nu^\Phi},\Gamma) \simeq \lim_{q\to\infty}\left(\frac{-1}{16\pi^2}\int_0^{2\pi}\Delta_b \mathrm{E}_{in}\left(4\pi^2\tau q^2\widetilde{D}(q_0,\frac{\pi}{2},\nu)\right)d\nu + \frac{1}{4\pi}\right) \\ &= \frac{-1}{16\pi^2}\int_0^{2\pi}\lim_{q\to\infty}\Delta_b\left(\mathrm{E}_{in}\left(4\pi^2\tau q^2\widetilde{D}(q_0,\frac{\pi}{2},\nu)\right)\right)d\nu + \frac{1}{4\pi} \\ &= \frac{-1}{16\pi^2}\int_0^{2\pi}\lim_{q\to\infty}\Delta_b\left(-\log\left(4\pi^2\tau\widetilde{D}(q_0,\frac{\pi}{2},\nu)\right) - \log(q^2) - \gamma\right)d\nu + \frac{1}{4\pi} \\ &= \frac{1}{16\pi^2}\int_0^{2\pi}\Delta_b\log\left(-\log\left(\widetilde{E}(q_0,\frac{\pi}{2},\nu)\right)\right)d\nu + \frac{1}{4\pi}, \end{aligned} \tag{7.28}$$

where γ is the Euler–Mascheroni constant. The asymptotic behavior of E_{in} for large q is studied in Appendix F. The estimator in eq. (7.28) is very similar to the DOT proposed in [Öza06], since both of them need to assume a constant ADC for all q. However, the one here introduced is an estimator of actual marginal probabilities, while the DOT is not. Besides, the numeric implementation of this estimator, described in Appendix D, is simpler than that of the DOT.

---

[1] A similar result to that in eq. (7.27) has been developed in [CR09], in parallel with the work here presented.

Name	Implementation	Function	Constant ADC	Radial integral error
Q–Balls	$\mathscr{G}\{E(\mathbf{q})\}(\mathbf{r})$	$\Psi(\mathbf{r})$	No	High (FRT)
DOT	—	$\Upsilon(\mathbf{r})$	Global	No
OPDT	$\frac{-1}{4\pi^2}\mathscr{G}\{\Delta E(\mathbf{q})\}(\mathbf{r})$	$\Phi(\mathbf{r})$	Local	High (FRT)
cQ–Balls	$\frac{-q_0^2}{4}\mathscr{G}\left\{\frac{1-E(\mathbf{q})}{\log E(\mathbf{q})}\right\}(\mathbf{r})$	$\Psi(\mathbf{r})$	Local	Low (disk)
cOPDT	$\frac{-1}{16\pi^2}\mathscr{G}\{\Delta_b \mathrm{E}_{in}(-\log E(\mathbf{q}))\}(\mathbf{r})+\frac{1}{4\pi}$	$\Phi(\mathbf{r})$	Local	Low (disk)
pQ–Balls	$\frac{-q_0^2}{4}\mathscr{G}\left\{(\log E(\mathbf{q}))^{-1}\right\}(\mathbf{r})$	$\Psi(\mathbf{r})$	Global	No
pOPDT	$\frac{1}{16\pi^2}\mathscr{G}\{\Delta_b \log(-\log E(\mathbf{q}))\}(\mathbf{r})+\frac{1}{4\pi}$	$\Phi(\mathbf{r})$	Global	No

Table 7.1: Summary of the estimators presented in Chapters 6 and 7 (bold), together with some other related estimators of interest. The implementation of the DOT in [Öza06] is not given since it does not admit a simple, closed–form expression. The optimal estimator would be that estimating $\Phi(\mathbf{r})$ (true probabilistic information), without making any assumption on the ADC, and without committing any error in the radial integral.

7.6 Relation with the FRT. Practical implementation

Four novel estimators have been introduced in this Chapter. They are described in eqs. (7.18), (7.25), (7.27) and (7.28). The first two are based on the circulation of a vector field along the boundary Γ of the domain Ω, so they will be referred to as cQ–Balls and cOPDT, where ‘c’ stands for ‘circulation’. The other two estimators are based on the integration of a given function on the whole orthogonal plane Π to the direction of interest $\mathbf{r}$, and will be referred to as pQ–Balls and pOPDT, where now ‘p’ stands for ‘plane’. Comparing the final expressions in the aforementioned equations, all the estimators are computed as the integral along Γ of a function which is related to the attenuation signal $E(\mathbf{q})$ in a relatively simple way. Consequently, once this related function is computed, the final estimator may be easily computed by means of the FRT as described in Section 6.3. Additionally, if the function is expressed in terms of SH expansions, the efficient numeric implementation proposed in [Des07] may be used (see Appendix D for details). The precise expressions for each estimator, together with a comparative analysis of the assumptions they require, and the orientation information they estimate, are given in Table 7.1.

All the estimators developed in this Chapter can be written as the FRT of a function of $E(\mathbf{q})$ (or the angular Laplacian of $E(\mathbf{q})$, for the estimators of $\Phi(\mathbf{r})$) that changes its angular contrast; for example, pQ–Balls may be seen as the conventional Q–Balls approach applied to a preconditioned signal computed from $E(\mathbf{q})$ as $-1/\log E(\mathbf{q})$. The corresponding contrast–enhancing functions for the remaining estimators are depicted in Fig. 7.2. Obviously, the domain of these functions is the interval $[0,1]$, since $E(\mathbf{q})$ is an attenuation of the baseline image. Note the different behavior of the functions for the estimators of $\Psi(\mathbf{r})$, monotonically increasing, and the estimators of $\Phi(\mathbf{r})$, monotonically decreasing. Besides, the functions for the estimators based on circulations are bounded, which ensures a certain stability in

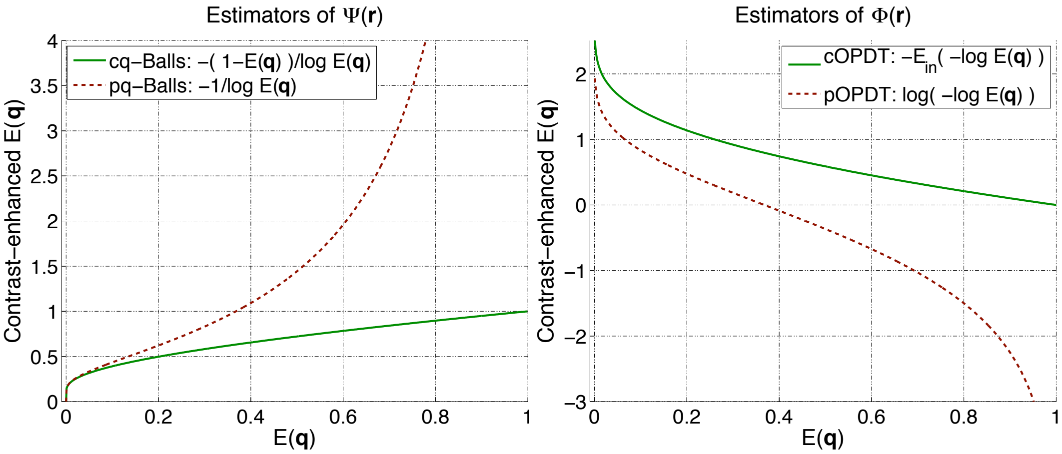


Figure 7.2: Contrast–enhancing functions applied to the attenuation signal $E(\mathbf{q})$ before the FRT is computed, for the estimators presented in this Chapter. Left: estimators of $\Psi(\mathbf{r})$. Right: estimators of $\Phi(\mathbf{r})$. In the latter case, the Laplace–Beltrami operator of the resulting signal has to be calculated prior to the computation of the FRT.

the computation of the orientation information. This is not the case for the functions corresponding to the estimators based on a constant ADC (or alternatively, based on the integration over the whole plane Π), which are unbounded.

7.7 Numerical simulations

The methodology for numerical validation to be carried out is the same as in Chapter 6. For a detailed review about the simulation parameters, variables of interest and possible scenarios, the reader is addressed to Section 6.6.1. Concerning the new estimation techniques introduced in this Chapter, they do not require any additional parameter apart from those inherited from their counterparts (regular Q–Balls or the OPDT). Therefore, the tuning of parameters is exactly the same as in the previous Chapter.

In the subsequent study, not only the four novel estimators presented are going to be tested. For comparison purposes, all the estimators reviewed in Table 7.1 are included in the analysis. To better organize such large amount of information, the results presented have been split in two groups: on one hand, those related with the estimation of the radial projection $\Psi(\mathbf{r})$; on the other, the ones related to the estimation of the true probability density (OPDF) $\Phi(\mathbf{r})$. Regarding the DOT, it is intended to estimate the probability profile at a given distance R_0, so–called $\Upsilon(\mathbf{r})$ in the previous Chapter. Since the performance of the DOT in terms of detection capability and resolution accuracy is closer to that of the OPDT than it is to Q–Balls, the results for $\Upsilon(\mathbf{r})$ are grouped in the second category.

7.7.1 Capability of resolving two crossing fibers

ODF estimators

To begin with, Fig. 7.3 illustrates the angular error in the detected fibers compared to the ground–truth directions, as a function of the initial angle between them. As it can be seen, all the estimators yield a very similar performance, as previously reported in [TV09b]. Nevertheless, pQ–Balls shows a greater error for

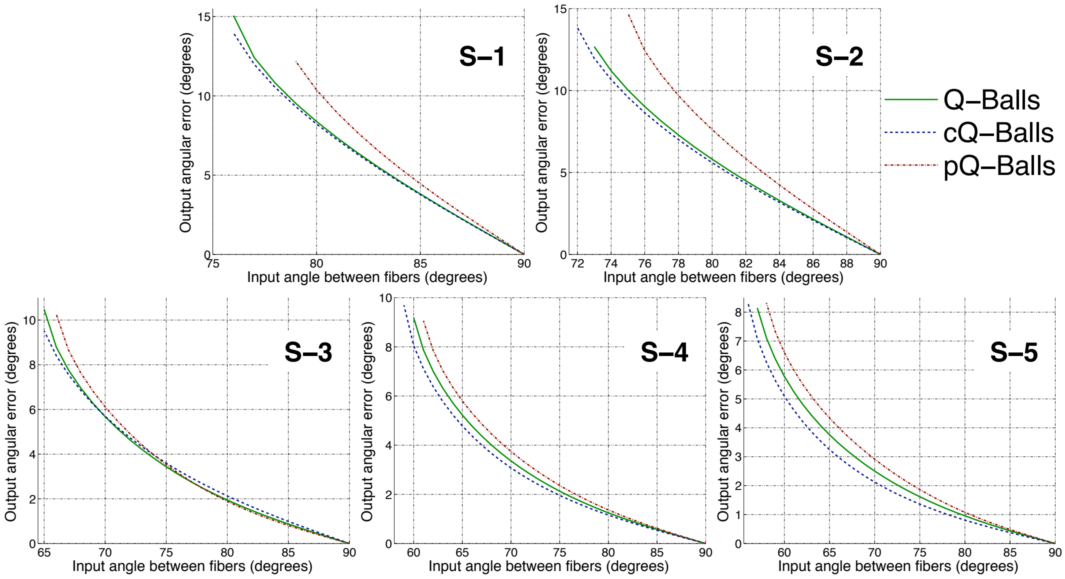


Figure 7.3: Angular errors in the estimation of two crossing fibers, as a function of the initial angle between the ground–truth directions, for the estimators of the ODF, $\Psi(\mathbf{r})$. Together with Q–Balls, the results for two novel estimators, cQ–Balls and pQ–Balls, are presented.

low b–values, and yet its resolution capability is poorer than that of the other two estimators: it fails to correctly recover the two crossing fibers for larger angles than Q–Balls or cQ–Balls do. For higher b–values, the resolution capability is practically the same for all the estimators, although the error committed by pQ–Balls is still slightly higher. Comparing cQ–Balls with Q–Balls, the former performs better practically in all situations. Although its advantage is only slightly noticeable, cQ–Balls shows a better resolution capability and lower detection errors. Fig. 7.4 shows glyph representations of the three estimators for scenario **S–4**. The differ-

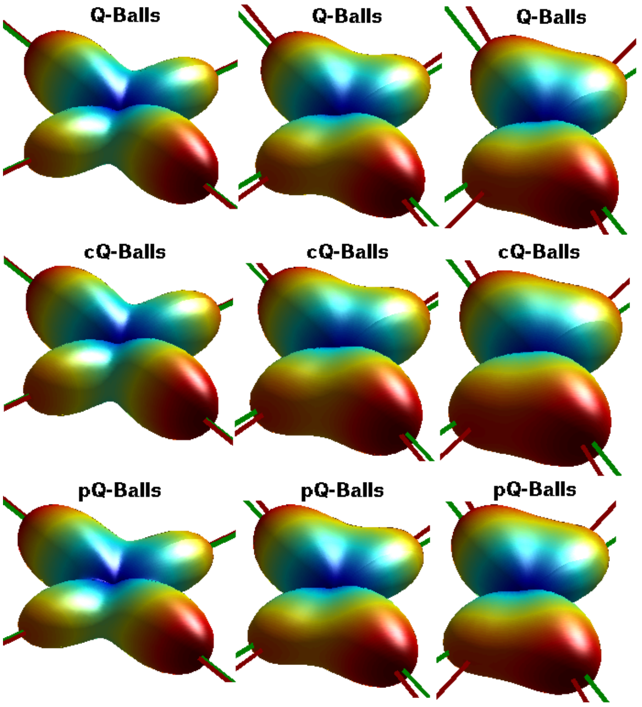


Figure 7.4: Glyph representations of the estimators of $\Psi(\mathbf{r})$ in a noise–free environment (**S–4**), for two fibers crossing in an angle of: 90^o (left), 80^o (center) and 70^o (right).

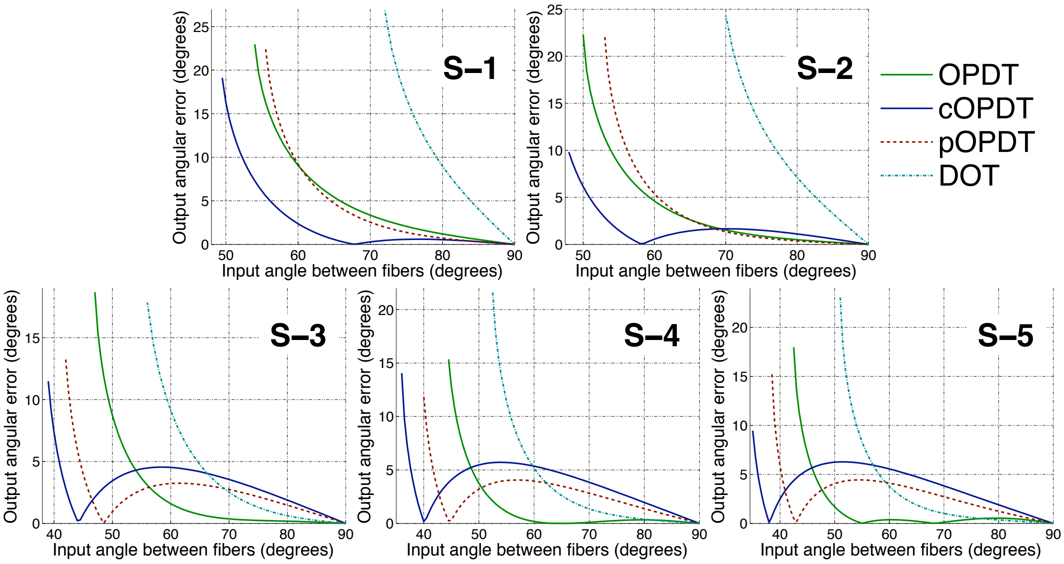


Figure 7.5: Angular errors in the estimation of two crossing fibers, as a function of the initial angle between the ground–truth directions, for the estimators of the OPDF $\Phi(\mathbf{r})$. Together with the OPDT, two novel estimators, the cOPDT and the pOPDT are tested. Additionally, the DOT is included for comparison purposes, though it is an estimator of $\Upsilon(\mathbf{r})$ and not $\Phi(\mathbf{r})$.

ences in the shapes of the estimations are completely negligible, corresponding to the behavior observed in Fig. 7.3. This first experiment allows to conclude that the error introduced by the computation of the FRT is in fact not so important: the reduction in the burring of the radial error achieved by cQ–Balls with the integration inside the disk Ω is almost compensated by the need to assume a locally constant ADC, so that the overall advantage is only marginal. On the other hand, pQ–Balls allows to completely remove the radial blurring, but at the expense of imposing a much more restrictive model, i.e. a constant ADC for all q; the performance of pQ–Balls is noticeably worsened for this reason, especially for lower b–values. Therefore, at least for the estimation of $\Psi(\mathbf{r})$, it follows that the error due to the FRT blurring is comparable to the error in the assumption of a slowly varying ADC (see Section 6.6.5), but much lower than the error due to the assumption of an absolutely constant ADC.

OPDF estimators

The counterpart of Fig. 7.3 for the estimators of the OPDF is shown in Fig. 7.5. In this case, the results are more difficult to interpret. For low b–values, the advantage of the cOPDT is clear. The resolution capability is drastically improved, and the detection errors, especially with few gradient directions, are as well reduced. For scenario **S–2**, however, the OPDT and the pOPDT may be preferable to resolve fiber crossings in larger angles; now, the pOPDT based on the assumption of a constant ADC yields a very similar behavior to the OPDT (contrary to Q–Balls, the OPDT needs a weak assumption on the ADC, so the pOPDT introduces a minor change in the model with respect to the OPDT than pQ–Balls does with respect to Q–Balls). Anyway, the advantage of any of the OPDF estimators over the DOT for low b–values remains clear. The pOPDT and the DOT share the same model for the attenuation signal and almost the same numerical implementation, so once again it becomes evident that the different behavior comes from the different orientation functions estimated. In particular, the importance of estimating true probabilistic information is demonstrated by this comparison. For larger b–values, the situation is different. First, both the pOPDT and the cOPDT show a very well differentiated

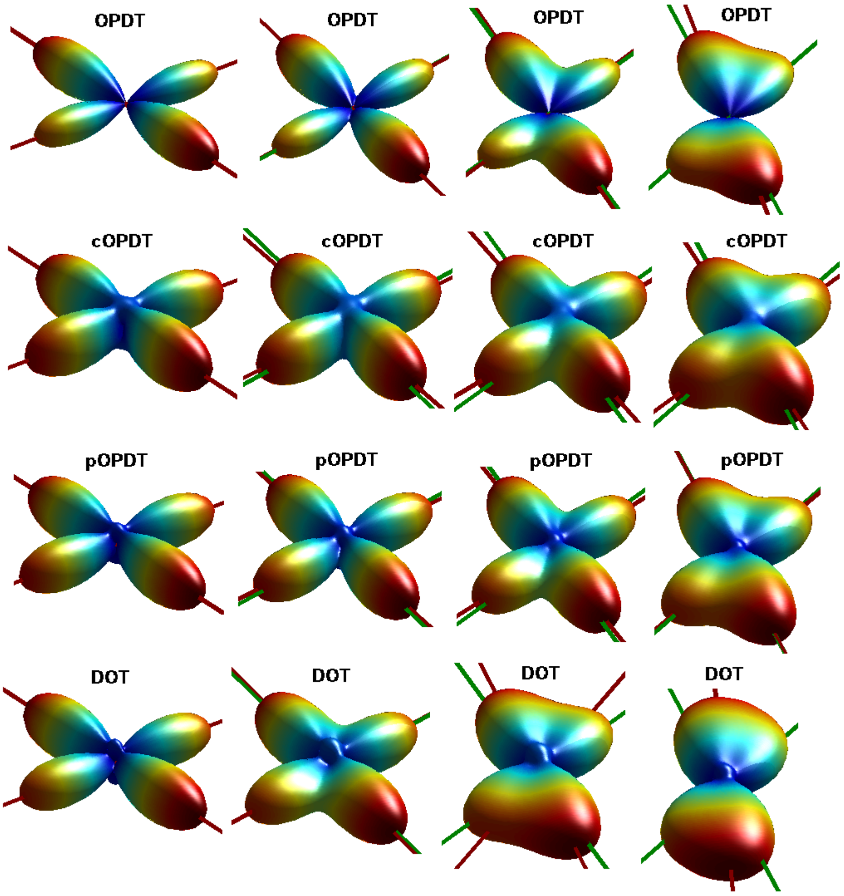


Figure 7.6: Glyph representations of the estimators of $\Phi(\mathbf{r})$ (and $\Upsilon(\mathbf{r})$, in the last row) in a noise–free environment (**S–4**), for two fibers crossing in an angle of, from left to right: 90^o, 70^o, 55^o and 47^o.

local minimum for certain input crossing angles. This behavior was observed with the OPDT in the previous Chapter, and explained in terms of the error in the estimation of the radial Laplacian: for low b–values, the inability to distinguish both fibers hinders this effect, so the local minimum is not appreciated. For high b–values, the resolution capability increases, and the effect of the estimation of the Laplacian conditions the behavior of the error. With the new estimators, the effect for high b–values is even more obtrusive. Yet, the peak for the cOPDT is more noticeable, appearing even for low b–values, contrary to the other estimators compared. As shown in Section 7.7.3, the possible origin of this artifact is the error committed when approximating the integral in the plane Π by the integral in the disk Ω; it will be shown that the error (and, what is more important, the variability of this error among all possible orientations) is reduced when the orientations of both fibers get closer. This effect, together with the increase of the accuracy of the locally constant ADC model, could explain the fast decay of the angular error before the estimator is finally unable to resolve the fiber crossing. Comparing the performance of the estimators, the cOPDT is still the one resolving lower crossing angles, followed by the pOPDT. However, for larger crossing angles, the OPDT is preferable due to its higher accuracy. Even the DOT can yield better results for nearly orthogonal fiber crossings. Nevertheless, neither the cOPDT nor the pOPDT show the theoretical limitation of the OPDT for arbitrarily high b–values, so they are still an interesting choice. In particular, the pOPDT may be a good trade–off between the resolution capability of the cOPDT and the accuracy of the OPDT. To finish this discussion, Fig. 7.6 shows glyph representations of the estimators compared in this Section. The most remarkable fact is that the cOPDT yields well

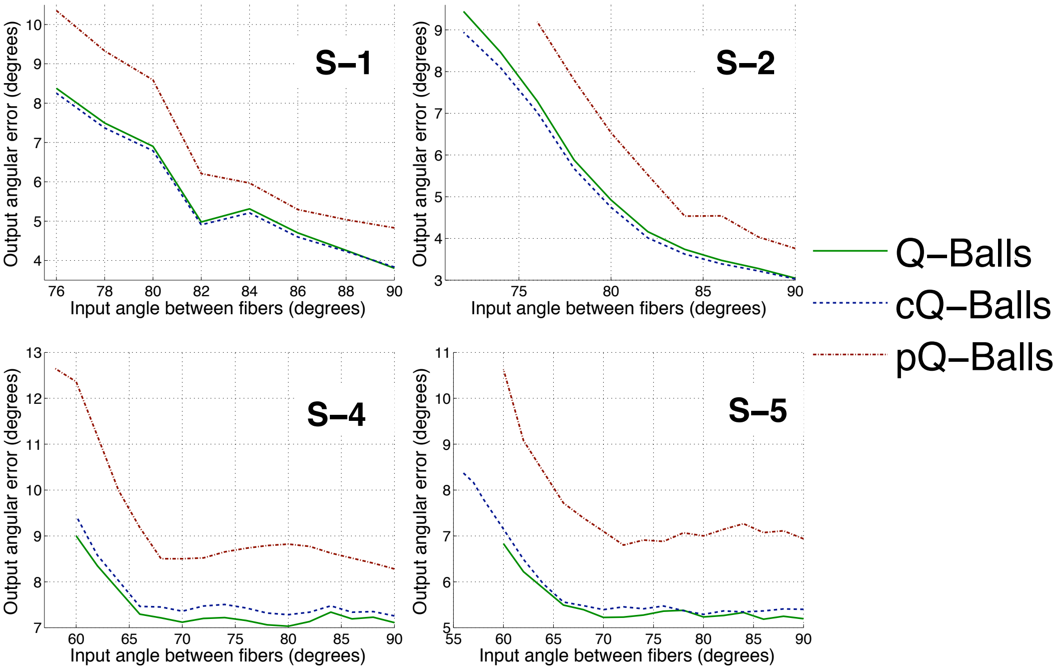


Figure 7.7: Angular errors in the estimation of two crossing fibers, as a function of the initial angle between the ground–truth directions, for the estimators of the ODF, $\Psi(\mathbf{r})$. Top: PSNR=13.33. Bottom: PSNR=5.

defined lobes even for fibers crossing in very small angles.

7.7.2 Behavior in the presence of noise

ODF estimators

Fig. 7.7 shows the angular errors in a noisy environment. Like in the previous Chapter, the PSNR has been fixed to PSNR=13.33 for scenarios **S–1** and **S–2** and PSNR=5 for **S–4** and **S–5**. The first conclusion is that Q–Balls and cQ–Balls behave practically the same. Although cQ–Balls is slightly more accurate for low b–values, Q–Balls seems to perform better for higher b–values (although for **S–5** cQ–Balls seems to have a better resolution capability). But concerning pQ–Balls, it is clearly less robust to noise, with much higher angular errors and, in general, a worse capability to resolve crossing fibers (except for scenario **S–4**; however, this result seems not very conclusive). To give a deeper insight into the effect of noise for the estimators of the ODF, Fig. 7.8 shows glyph representations for **S–2**, for an angle of 90^o and PSNR=13.3. While the shapes of the ODF recovered by both Q–Balls and cQ–Balls are very similar, pQ–Balls clearly fails to recover the orientation information, yielding an inadequate estimation. The conclusion is that pQ–Balls is far more sensitive to noise than the other estimators; even in a noise–free environment (see Fig. 7.3) it yields worse estimates than Q–Balls or cQ–Balls,

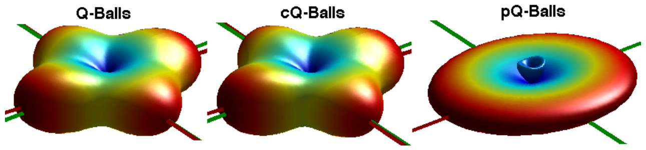


Figure 7.8: Glyph representations of the estimators of the ODF in a noisy scenario: **S–2** with PSNR=13.3. Ground–truth angle between the fibers is 90^o.

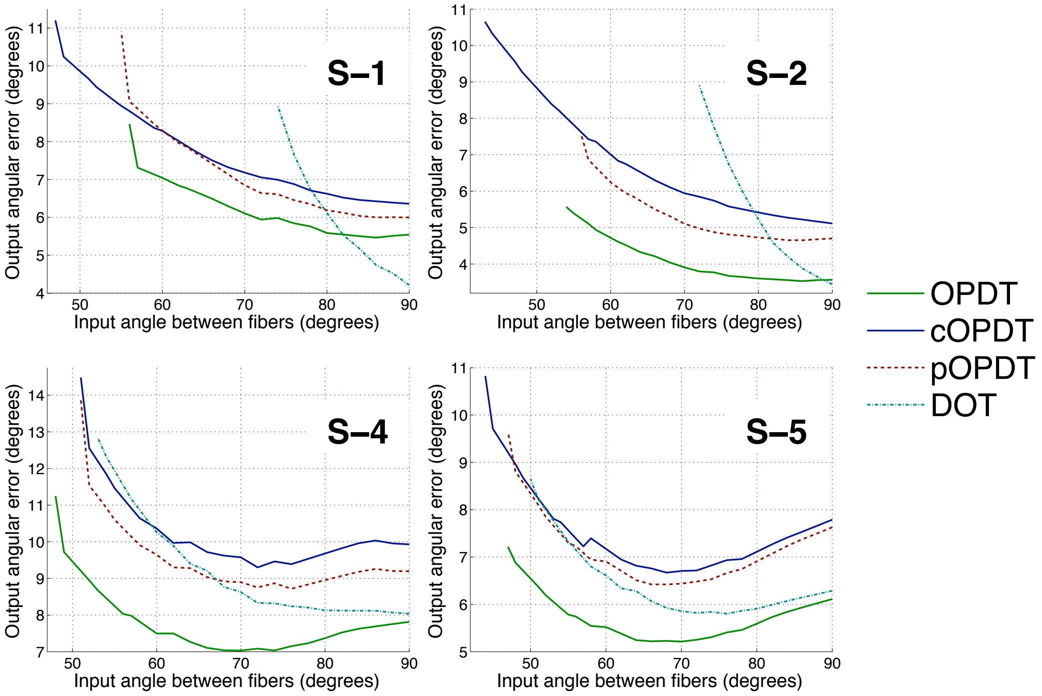


Figure 7.9: Angular errors in the estimation of two crossing fibers, as a function of the initial angle between the ground–truth directions, for the estimators of the OPDF, $\Phi(\mathbf{r})$. Top: PSNR=13.33. Bottom: PSNR=5.

so this other estimator is not adequate.

OPDF estimators

As always, Fig. 7.9 is the counterpart for the corresponding estimators of the ODF in Fig. 7.7. The obvious conclusion is that the OPDT is more robust to noise than both the cOPDT and the pOPDT. But for low b–values (and for scenario **S–5**), the cOPDT has still an important advantage recovering small crossing angles. For high b–values, the cOPDT and the pOPDT have a similar behavior (the pOPDT yields more accurate estimates but the cOPDT has a better resolution capability): the estimation accuracy is worsened so that it gets close to that of the DOT, losing in part the advantage of the OPDT over this estimator. Note, however, that the DOT fails to recover the fiber crossing before the cOPDT and the pOPDT do, so they are still preferable. In these cases, the OPDT shows a very similar resolution capability, but it is notably more accurate. Generally speaking, FRT–based estimators are more robust to noise than their 'circulation' and 'plain' counterparts. This observation may be explained by the averaging inherent to the FRT: the integration inside the Bessel–shaped tubes blurs the orientation information yielding poorer estimates, but at the same time averages the radial information, partially palliating the effect of noise. Under a different point of view, this phenomenon may be clarified considering the curves in Fig. 7.2: the enhancement of the contrast of $E(\mathbf{q})$ carried out with the 'circulation' and 'plane' approaches may somehow amplify the noise in the attenuation signal, aggravating its effect. Finally, Fig. 7.10 shows glyph representations for scenario **S–2** and PSNR=13.33. Although not so evident as it is for pQ–Balls, the higher sensitivity to noise of the pOPDT is highly noticeable. This estimator yields unnatural shapes for the OPDF, with flattened surfaces and a large blurring (and thus a great uncertainty) in the direction of local maxima. Comparing the OPDT and the cOPDT, the former is almost unaltered by the presence of such large amount of noise, while the latter shows a certain

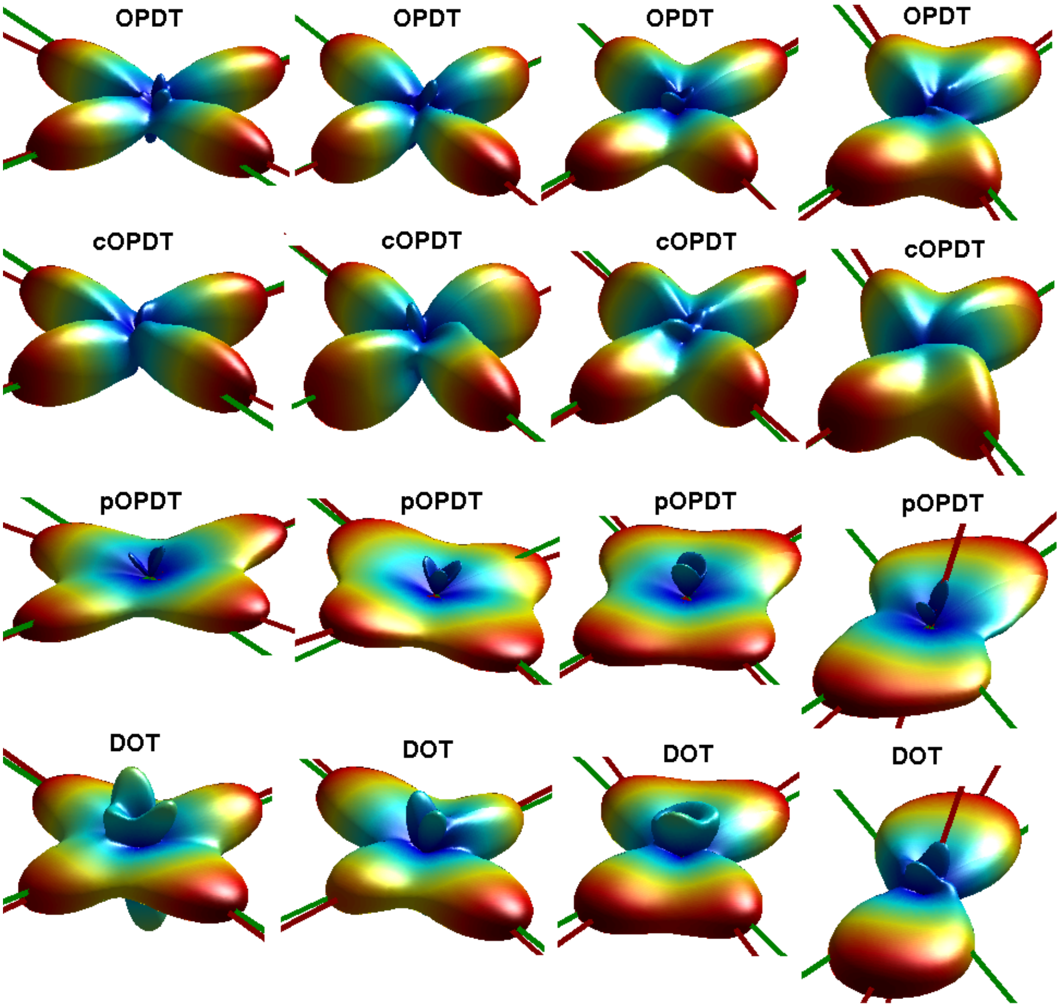


Figure 7.10: Glyph representations of the estimators of the OPDF in a noisy scenario: **S–2** with PSNR=13.3. Ground–truth angles between the fibers are, from left to right: 90^o, 80^o, 70^o and 60^o.

distortion in the recovered OPDF. This is consistent with the higher angular errors measured for the cOPDT in Fig. 7.9. As a final remark, all these estimators are able to give better approximations than the DOT.

7.7.3 Accuracy of the integrals inside the disk Ω

As a final experiment, the accuracy in the computation of the radial integrals in terms of integrals inside Ω is now tested. To that end, the same tensor configuration as in the previous experiments has been used. For each input angle between the crossing fibers, 51 gradient directions uniformly distributed in the surface of the sphere have been chosen. For each of them, the integral of the Laplacian of the attenuation signal, computed with eq. (E.3) in Appendix E, has been computed in the whole plain Π (this integral is denoted I_p) and inside the disk Ω (denoted I_c) for a range of values of q_0. The relative error is defined as:

$$\varepsilon = \frac{I_p - I_c}{I_p} > 0, \tag{7.29}$$

where I_p is obviously negative since its (normalized by $1/4\pi^2$) opposite is a probability density. The most representative results are depicted in Fig. 7.11. The first

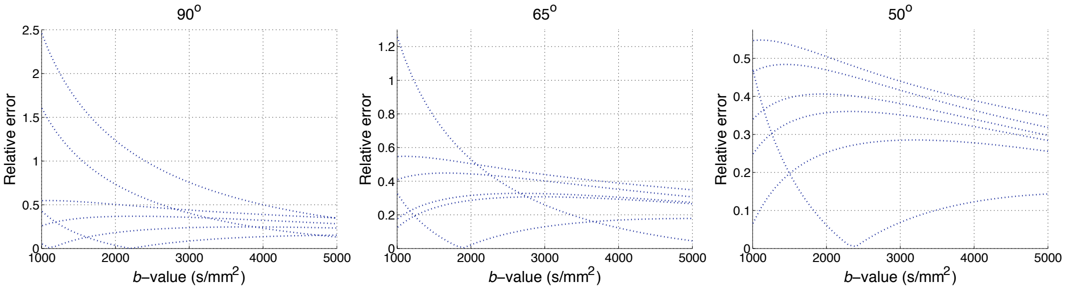


Figure 7.11: Relative errors in the estimation of the integral in the plane Π as the integral in the disk Ω, as a function of the b–value, for crossing angles of 90^o (left), 65^o (center) and 50^o right. To keep the interpretation of the figure as clear as possible, only the errors for representative directions are shown. Of course, these directions include those yielding maximum and minimum errors.

comment is that the relative errors are quite large. Even so, the disk approximation is far more accurate than the approximation as the integral along the boundary of Ω, and it avoids the problem with the negativity of the orientation functions. It is worth noticing that the relative error in the integration decreases as the directions of the fiber bundles get closer (compare the scale range of the three figures). It is even more important to note that the relative errors for different directions become similar for higher b–values and closer fiber directions. If the relative errors are similar for all directions, it means that the integrals for all possible orientations are underestimated in the same extent. On the contrary, if the variability of this error is larger, the integrals for certain directions will be underestimated more than for others, which introduces an additional distortion in the computation of the OPDF. As mentioned above, this is an additional possible explanation for the anomalous decay of the angular error of the cOPDT in Fig. 7.5.

However, the main reason for this behavior has to rely on the accuracy of the model for the attenuation signal. This same artifact can be observed in the curves corresponding to the pOPDT, even when the integration for this estimator is exact, and thus this source of error is completely removed. The local decay of the error is not so noticeable as with the cOPDT, but it is still far more important than with the OPDT. For the pOPDT, it is necessary to assume a global model for the attenuation signal, as opposed to the local assumption made with the OPDT: the smaller the angle between the fiber bundles, the closer the model to a pure tensor configuration, and hence the more accurate the fitting of the mono–exponential decay. This reasoning explains why the error due to the supposition of a constant ADC decreases for small input angles. Besides, the premise is far more restrictive with the pOPDT than it is with the OPDT, and this artifact is more noticeable. For the cOPDT, the error in the approximation of the plane integral has to be added to the error originated by the model, so the anomalous behavior is even stressed.

7.8 Conclusion

The FRT has been proved an extremely valuable mathematical tool for HARDI estimators, beyond its direct use as an estimator of orientation information. This kind of information is computed as (weighted) line integrals of the diffusion propagator along the directions of interest, whose calculation has to be carried out in the dual Fourier domain (the **q**–space). It has been demonstrated that a line integral in the **R**–space is equivalent to an integral in the orthogonal plane of the **q**–space. Due to the symmetry of the problem, this plane has to be parametrized using polar coordinates: the FRT represents the integration on the angular coordinate of such system, which explains the ubiquitousness of this operator in the estimators presented in this Chapter.

With Q–Balls and the OPDT, the FRT is used directly as an estimator, so the radial integral in the aforementioned polar coordinates system is ignored. Obviously, this is an important source of error, but at the same time it allows to get rid of any assumption on the behavior of the attenuation signal (for Q–Balls) or at least of strong assumptions (for the OPDT). A number of techniques to minimize the effect of this source of error have been proposed in this Chapter, varying the modeling assumptions of $E(\mathbf{q})$. For the estimators of $\Psi(\mathbf{r})$ (based on Q–Balls), it has been shown that the error introduced by unrealistic models may be more important than the blurring due to the FRT: cQBalls is only marginally better than Q–Balls, and pQ–Balls performs worse and is more sensitive to noise. For the estimators of $\Phi(\mathbf{r})$, on the contrary, the advantage of reducing the FRT blurring has been evidenced. The novel cOPDT shows a better resolution capability and higher accuracy than the OPDT for low b–values; for high b–values, its accuracy is worse than that of the OPDT, but its resolution capability is still better. The cOPDT does not require additional assumptions with respect to the OPDT, so alleviating one of its sources of error may drive to better results. Yet, the pOPDT may perform similar to the OPDT for low b–values (for high b–values, it shows a better resolution capability, although its accuracy is poorer), even when the modeling of the attenuation signal for the pODPT is far more restrictive.

Generally speaking, there are two fundamental sources of error in the estimation: the error in the estimation of radial integrals (which is reduced with the cOPDT and completely removed with the pOPDT), and the error due to the incorrect modeling of the attenuation signal (which is more severe for the pOPDT; although the OPDT and the cOPDT share the same model, at the sight of Fig. 7.5 it seems that the integration of such model to apply Stokes' theorem is more sensitive to the error than its derivation to compute the radial Laplacian). The reduction of one of these errors has to be done at the expense of increasing the other, and therefore a trade–off has to be reached for each particular situation.

The new estimation techniques yield results which are less robust to noise than those obtained by the direct application of the FRT. This fact may be explained by the elimination of the FRT averaging or by the contrast–enhancement described in Fig. 7.2. Nonetheless, the noisy scenarios tested in the numerical simulations are extreme cases, with a very poor SNR. Moreover, the filtering techniques described in Chapter 5 may be used to drastically improve the quality of the HARDi data sets, so the new estimators proposed, mainly the cOPDT and the pOPDT, may indeed result advantageous in many situations.

In particular, the results in Fig. 7.5 suggest very useful guidelines to choose the appropriate estimator for a particular situation, assuming that the denoising step allows to consider a high enough SNR. These guidelines may be seen as an heuristic to determine the aforementioned trade–off between the contributions of each source of error. For low b–values, it seems clear that the cOPDT is the best choice. For high b–values, the OPDT should be used unless very small crossing angles need to be detected (this could be the case for particular regions of the brain like the cerebellar peduncle). In this case, the pOPDT may be more appropriate, but if a high enough SNR cannot be assumed, the cOPDT is preferable (see for example Fig. 7.10). Additionally, it is important to stress that neither the cOPDT nor the pOPDT show the theoretical limitation of the OPDT for arbitrarily high b–values, so they can be used regardless of the scanning parameters. Finally, Q-Balls-related estimators, and above all cQ-Balls, can be very interesting in case only large angles have to be detected, due to their higher robustness to noise.

As a final remark, the usefulness of the estimation of true probabilistic information, firstly discussed in Chapter 6, remains evident from the results in this Chapter. In general terms, the pOPDT shows an overall worse accuracy than the OPDT and the cOPDT, but yet it is more accurate and has a better resolution capability than the DOT. Except for very wide crossing angles, its robustness to noise is also higher. The modeling of the attenuation signal, based on a constant ADC for all q, is exactly the same for the pOPDT and the DOT. Their numerical implementations, based on SH expansions and analytical integration, are very similar as well. Consequently, the reason for their different performances has to rely on the different orientation information estimated by each of them: only the pOPDT estimates a true marginal probability density.

8

Illustrative examples on fiber populations estimates

This Chapter shows qualitative results of fiber orientations estimators over real data sets. The aim is to compare the estimators presented in the previous Chapters (and some of the previously introduced in the related literature), in different scenarios. Besides, the advantages of filtering DWI volumes prior to the estimation of fiber populations are evidenced. To that end, relevant Regions of Interest (ROI) have been selected where fiber crossings and complex micro–architectures are especially noticeable. At each voxel of the corresponding ROI, a glyph standing for the estimated orientation information given by the corresponding technique is represented. The criteria to decide whether or not an estimation is appropriate is the agreement of the visual results with known neural architectures, as described in [Mor05]. Analogous representations to those in that work, including labels for the principal tracts present at each ROI, are given for guidance.

8.1 Introduction

Each technique on the Diffusion Imaging framework presented has been quantitatively evaluated throughout this dissertation. However, the numerical validation compels to use synthetic data which is not always representative of real scenarios. With respect to the DWI filtering techniques, it was justified in Chapter 5 that the phantom data used had a certain level of realism which allowed to infer general enough conclusions. Moreover, extensive experiments with real data were carried out over very different data sets. The same can be said about the statistical characterization proposed in Chapter 3, for which the assumptions on the models used were tested against real–world examples of scanned data. In Chapter 4, the study was purely theoretical: the results of interest were statistical parameters (the bias and the variance) which could be analytically computed given the SNR, the number of coils, and the configuration of gradient directions in the scanner. The validation in this case is intrinsically different; compared to filtering techniques, in which the behavior for phantom data in given scenarios had to be generalized to a realistic situation, the analysis carried out in Chapter 4 yields a closed form, ground truth solution for each SNR, tensor configuration, and number of gradients, so no generalization is required.

Thus, only the results in Chapters 6 and 7 lack an intensive validation with real data. For these techniques, the way to intend acceptably realistic phantoms is not evident, so a study with merely synthetic data had to be accomplished. Besides, each part of the framework has been validated separately, but the impact of DWI filtering on the estimation of fiber populations, for example, has been only superficially explored.

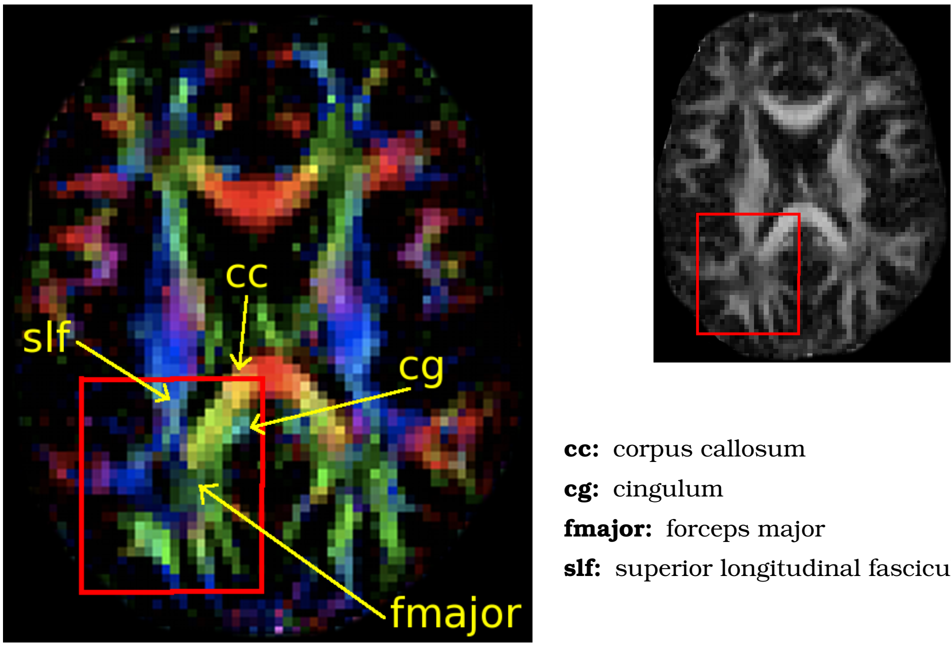
The aim of the present Chapter is to overcome these two illnesses. To that end, the estimation techniques reviewed in Table 7.1 are compared over the real data sets presented in Section 1.4. As mentioned in Chapter 6, the parameters used for the experimental part (number of SH coefficients, value of λ, R_0 for the DOT, and so on) were fitted to a realistic situation, so they are going to be used for all the experiments presented. To represent the enormous amount of information provided by the estimation of fiber populations at each voxel, a glyph representation has been chosen (see Section 2.6.1). At each voxel, a three–dimensional surface is plotted in spherical coordinates. The distance of the surface to the origin (the center of the voxel) at each particular orientation represents the probability density (or any other corresponding orientation information) for such orientation. To facilitate the interpretation of this information, the surface is colored following color–coding conventions (see Section 2.6.1), and the glyph is represented over a background whose gray level stands for the Generalized Anisotropy (GA, see eq. (2.42)) of the corresponding voxel.

As well as testing the capability of the estimators to resolve complex architectures, it is interesting to test the impact of noise removal algorithms in the final result. Among the techniques described in Chapter 5, LMMSE–N seemed to outperform the others in many cases, being especially well suited for HARDI scenarios. Consequently, this is the technique used in this Chapter. A value of $N = 15$ has shown an adequate behavior for all volumes, so this is the number used. For the estimation windows, a neighborhood of $5 \times 5 \times 3$ voxels has been used in all cases. All volumes have been pre–processed with this algorithm except otherwise noted.

8.2 Volume CSIRO1

As stated in Section 1.4, this is a typical DWI volume for HARDI estimation, which has both a great number of gradient directions (60), and a large b–value (3000 mm/s^2), which results in a high angular contrast. As a consequence, all the estimators are expected to provide accurate descriptions of the underlying anatomies. As a major drawback, the large b–value induces a poor SNR compared to other DWI data sets dealt with in this dissertation. Besides, its spatial resolution is lower than that of other available volumes. To begin with, Fig. 8.1 shows an axial slice in the middle brain, where a number of tracts of interest cross inside the corresponding Region of Interest (ROI). Although the resolution of the image does not allow a very intuitive representation, the color coding allows to properly interpret each fiber tract: red structures (like the center of the corpus callosum) follow the direction of the 'x' axis (horizontal in the image); green structures (the forceps major) follow the 'y' axis (vertical in the image); blue structures (the superior longitudinal fasciculus) follow the 'z' axis (transverse to the plane of the image). Those structures following oblique orientations, such as the part of the corpus callosum inside the ROI, are represented as non–primary colors, such as yellow for the aforementioned structure. This same conventions are kept in the remainder of the present Chapter.

The ROI in Fig. 8.1 is analyzed in detail in Fig. 8.2 with Q–Balls, the DOT and the cOPDT. No additional OPDT–based estimators are tested, since their qualitative behavior is very similar and such comparison would not contribute any substantial information. As a first comment, note that all estimators are able to correctly account for the anatomies suggested in Fig. 8.1. Glyphs are oriented following the directions indicated by the color coding, and their orientations have



cc: corpus callosum

cg: cingulum

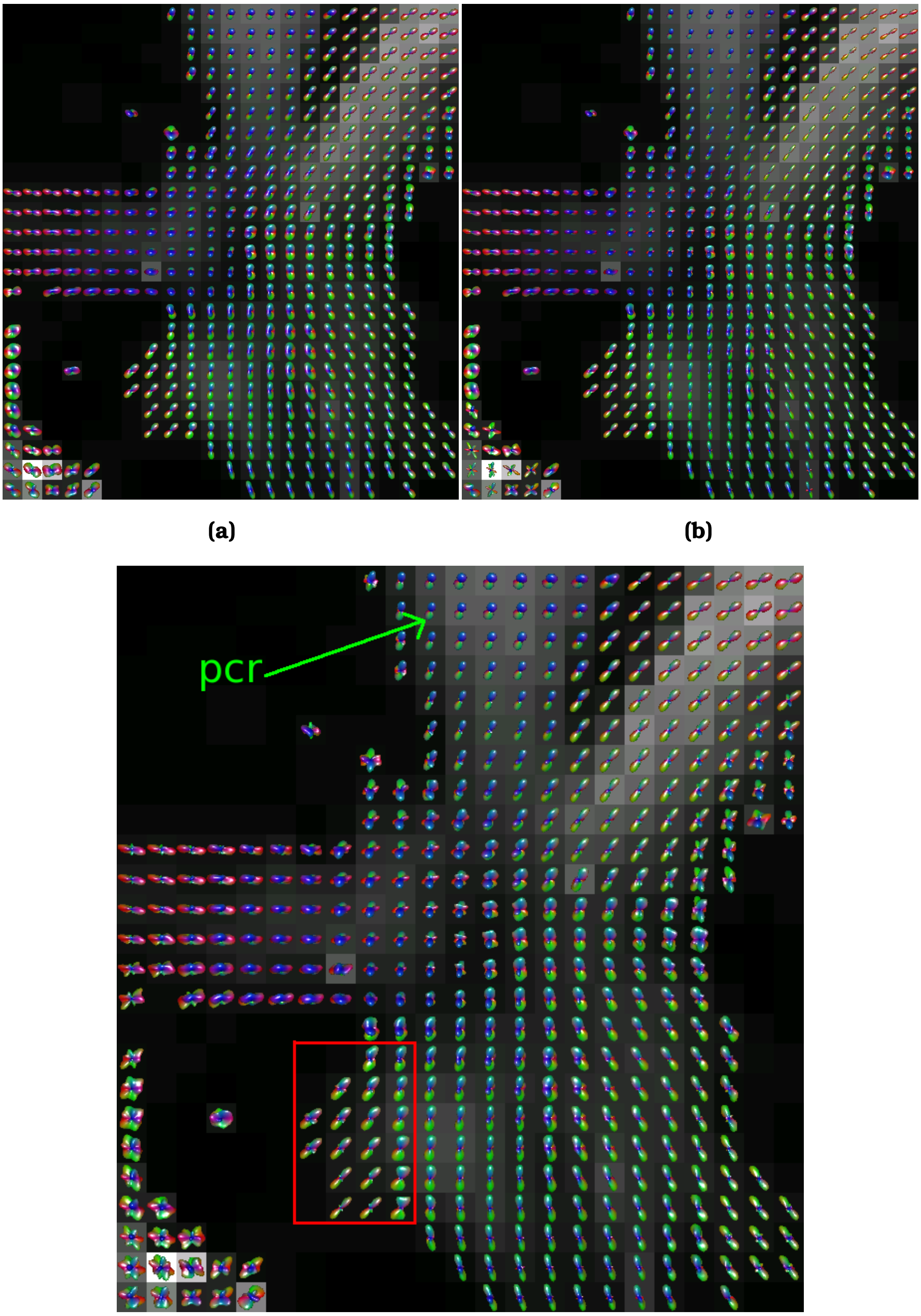
fmajor: forceps major

slf: superior longitudinal fasciculus

Figure 8.1: CSIRO1: axial slice in the middle brain (slice 27). Color–coding representation plus FA map. The ROI is marked in red together with some tracts of interest (yellow).

an appropriate spatial coherence which allows to interpret the information they provide. Moreover, HARDI estimation reveals fiber structures which remain hindered with the conventional DTI analysis. In particular, red structures entering the ROI on its left–hand side are almost hampered in Fig. 8.1 by the superior longitudinal fasciculus; in Fig. 8.2, it becomes evident that these structures cross the ROI and join the forceps major and the corpus callosum. All estimators are able to correcly resolve these crossings, which manifest on the left–hand side of the pictures in Fig. 8.2 with well defined cross–shaped glyphs. Moreover, the posterior corona radiata, which is not visible in Fig. 8.1, is also revealed by HARDI estimators. It corresponds to the longitudinal green structures following the 'y' axis in the top of the image. Although its presence is quite subtle, crossings with the blue structures corresponding to the superior longitudinal fasciculus are still visible.

Another important conclusion is that all the estimators compared show a very similar behavior, none of them yielding an overall better performance. This result should not be surprising, since the experiments over synthetic data sets carried out in previous Chapters have evidenced that, for large *b*–values, all HARDI estimators are able to resolve crossings even for rather small angles. In the top–right, the angle formed by the termination of the corpus callosum and the posterior corona radiata/superior longitudinal fasciculus is large enough, so the three approaches are able to distinguish between them. For very small crossing angles, however, it is expected that the cOPDT shows better resolution capabilities. For the sake of comparison, a new ROI has been selected in Fig. 8.2. Inside this window, the forceps major splits in two branches forming a small angle. As can be seen, the cOPDT resolves these crossings for most of the voxels inside this ROI; although the glyphs are not cross–shaped, and therefore the fiber directions are not so evident, two maxima can be identified in most of the voxels in the third and fourth columns of the window. This behavior is thoroughly analyzed in Fig. 8.3; the subtle maxima found in some glyphs obtained with the cOPDT, which in turn have not been detected by the other estimators, have been highlighted. According to the results presented with synthetic data, the cOPDT shows a better capabil-



(a) (b)

(c)

Figure 8.2: CSIRO1: detail of the slected ROI in axial slice 27, for (a) Q–Balls, (b) the DOT, and (c) the cOPDT. The third picture has been enlarged for illustration purposes. Glyphs are colored following color–coding conventions for each axis, and the LMMSE–15 has been used to pre–process the DWI.

ity to resolve fibers crossing in very small angles, which are otherwise completely blurred with Q–Balls and the DOT. Nevertheless, the cOPDT presents more spurious lobes, which is an undesired characteristic.

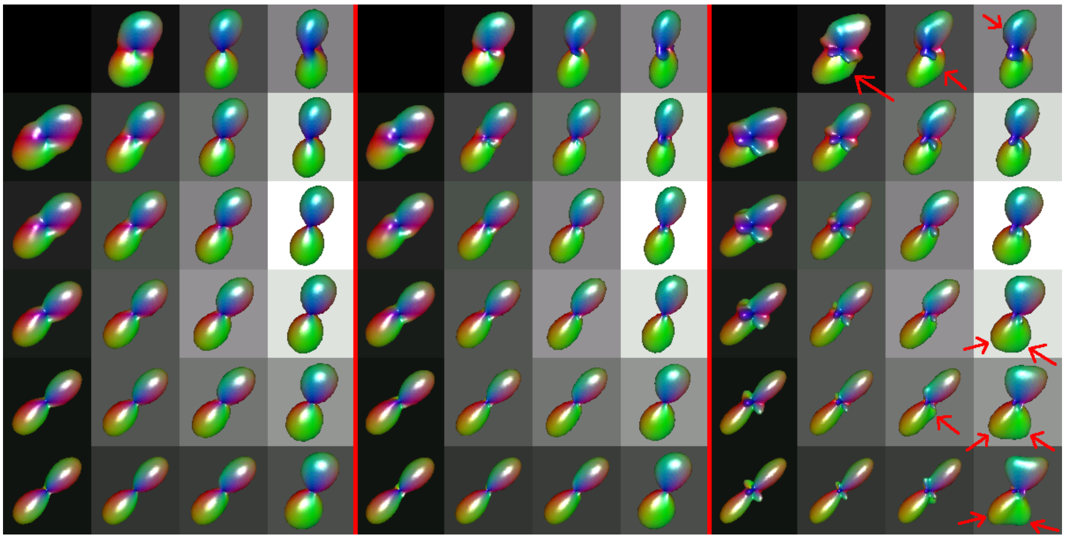


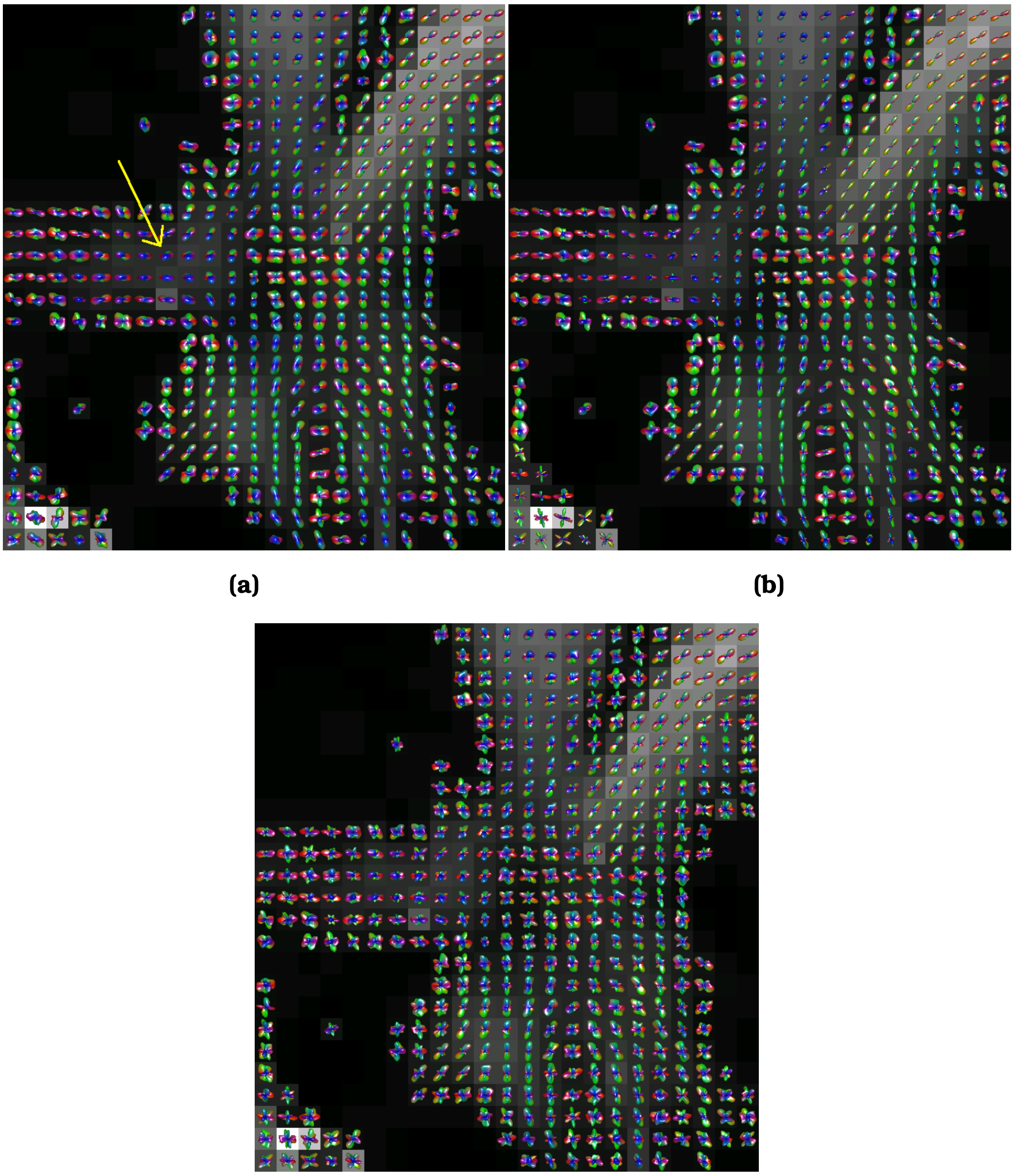
Figure 8.3: CSIRO1: Comparison of estimators in a small ROI inside axial slice 27. From left to right: Q–Balls, the DOT, and the cOPDT.

In the next experiment, the convenience of DWI filtering prior to the estimation of fiber populations is assessed. To that end, Fig. 8.4 presents analogous results to those in Fig. 8.2 computed over the original data set, without processing it with the LMMSE–15 technique. The first comment is obvious: the overall spatial coherence is drastically reduced in the absence of unbiased filtering for all estimators. The structures are still recognizable, mainly due to the use of color–coding, but the orientation of the glyphs is not always coherent with the predictable directions. Yet, the appearance of spurious lobes becomes more evident for the DOT and the cOPDT, while Q–Balls shows a more stable behavior with this respect, consistently with the results obtained in previous Chapters over synthetic data.

One interesting issue is that the red (transverse) structure coming from the left–hand side of the ROI is hindered at some places, if the image is not filtered, with the DOT and the cOPDT. Consider the region marked with a yellow arrow in the first picture of Fig. 8.4, where such structure crosses the superior longitudinal fasciculus. This crossing confuses with the spurious lobes created by the DOT and the cOPDT, so its interpretation is very difficult. Once the filter is applied, see Fig. 8.2, the structural information is correctly accounted, the spurious lobes are mostly eliminated, and the enhancement of the overall spatial coherence allows a very intuitive representation of the underlying neural architecture. A similar result was presented in Chapter 5 (see Fig. 5.12), as a first example of the potential of the filtering techniques presented.

It is also worth noticing that, when using Q–Balls, this crossing is still visible without LMMSE filtering. This reinforces the previous argument that this estimator is in general more robust to noise. Nevertheless, once the filter is applied, as shown in Fig 8.3, both the DOT and the cOPDT render this structure more visible than Q–Balls does, which is clearly an argument in favor of their use.

As a final discussion, analogous results to those in Fig. 8.3 are presented in Fig. 8.5 when no previous processing of the DWI is performed. All the previous comments apply to this new figure: first, undesired lobes not corresponding to any particular fiber architecture appear with the DOT and the cOPDT. Yet, this problem persists also for Q–Balls (see the first two columns of the corresponding picture), although once again this estimator is more robust to noise. Now, the splitting fibers inside the ROI are completely precluded with all the estimators, and the information provided by the glyphs is harder to interpret.



(a)

(b)

(c)

Figure 8.4: CSIRO1: detail of the slected ROI in axial slice 27, without LMMSE filtering, for (a) Q–Balls, (b) the DOT, and (c) the cOPDT. Glyphs are colored following color–coding conventions.

From the previous discussion, some conclusions should remain evident. Most of them are supported by the quantitative analysis carried out in Chapters 6 and 7. First, all the estimators tested (and, hopefully, any other reported in the related literature) are able to yield accurate estimates of fiber populations with DWI data sets specifically designed for this task; in particular, the b–value of data set CSIRO1 is large enough to provide an adequate angular contrast. As shown later on in this Chapter, this property is critical for some estimators. The second conclusion is related to the prior processing of DWI. It has been shown that the LMMSE–N technique not only improves the spatial coherence and produces smoother results, but it is also able to reveal structures whose visibility is dramatically reduced otherwise. This property is especially interesting in the case of OPDT–like estimators (and also the DOT), for which the appearance of spurious

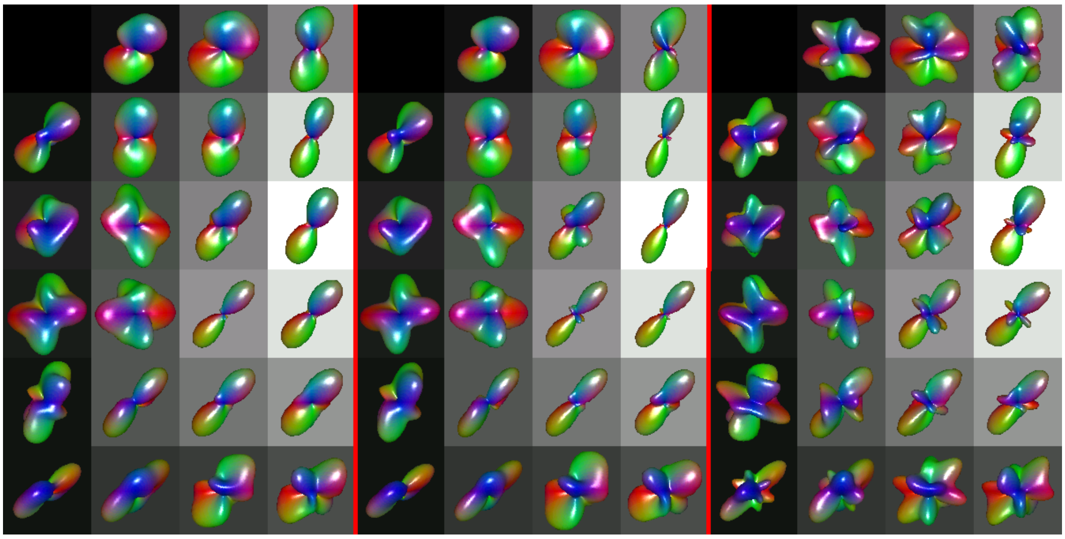
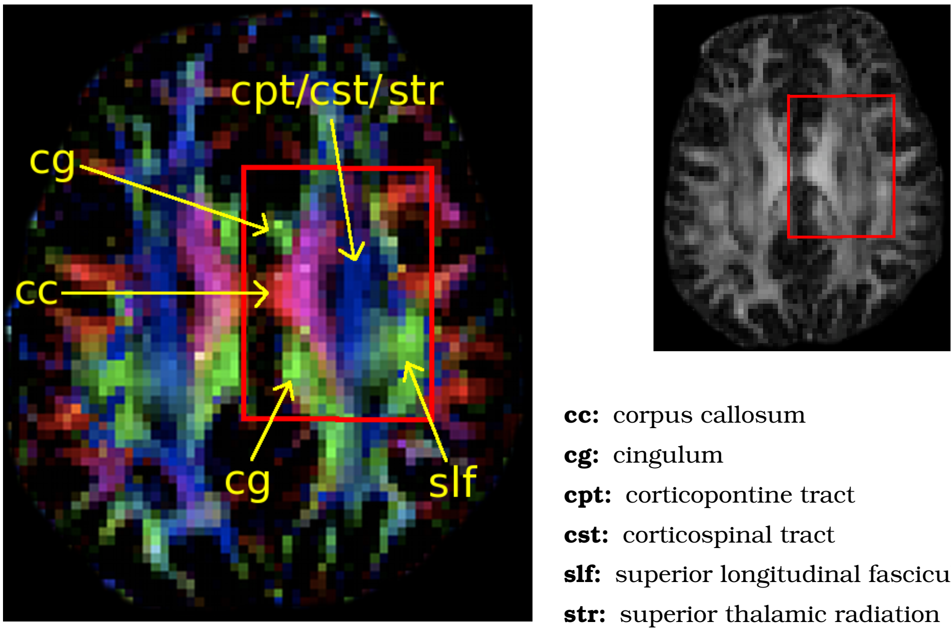


Figure 8.5: CSIRO1: Comparison of estimators in a small ROI inside axial slice 27, without LMMSE filtering. From left to right: Q–Balls, the DOT, and the cOPDT.

lobes can hinder the interpretation of the orientation information under some circumstances. Finally, the advantage of the cOPDT (and also the OPDT and the pOPDT, though the corresponding results have not been explicitly shown) for the discrimination of certain structures, such as fibers crossing in small angles, has been shown as well. The counterpart for this property is the presence of spurious lobes with the cOPDT, which can be more problematic than with the DOT in some situations.

To continue the discussion, Fig. 8.6 shows ans axial slice of CSIRO1 in the upper brain. The ROI selected is of special interest due to the great number of tracts of interest present inside it, following very different directions. Once again, the



cc: corpus callosum

cg: cingulum

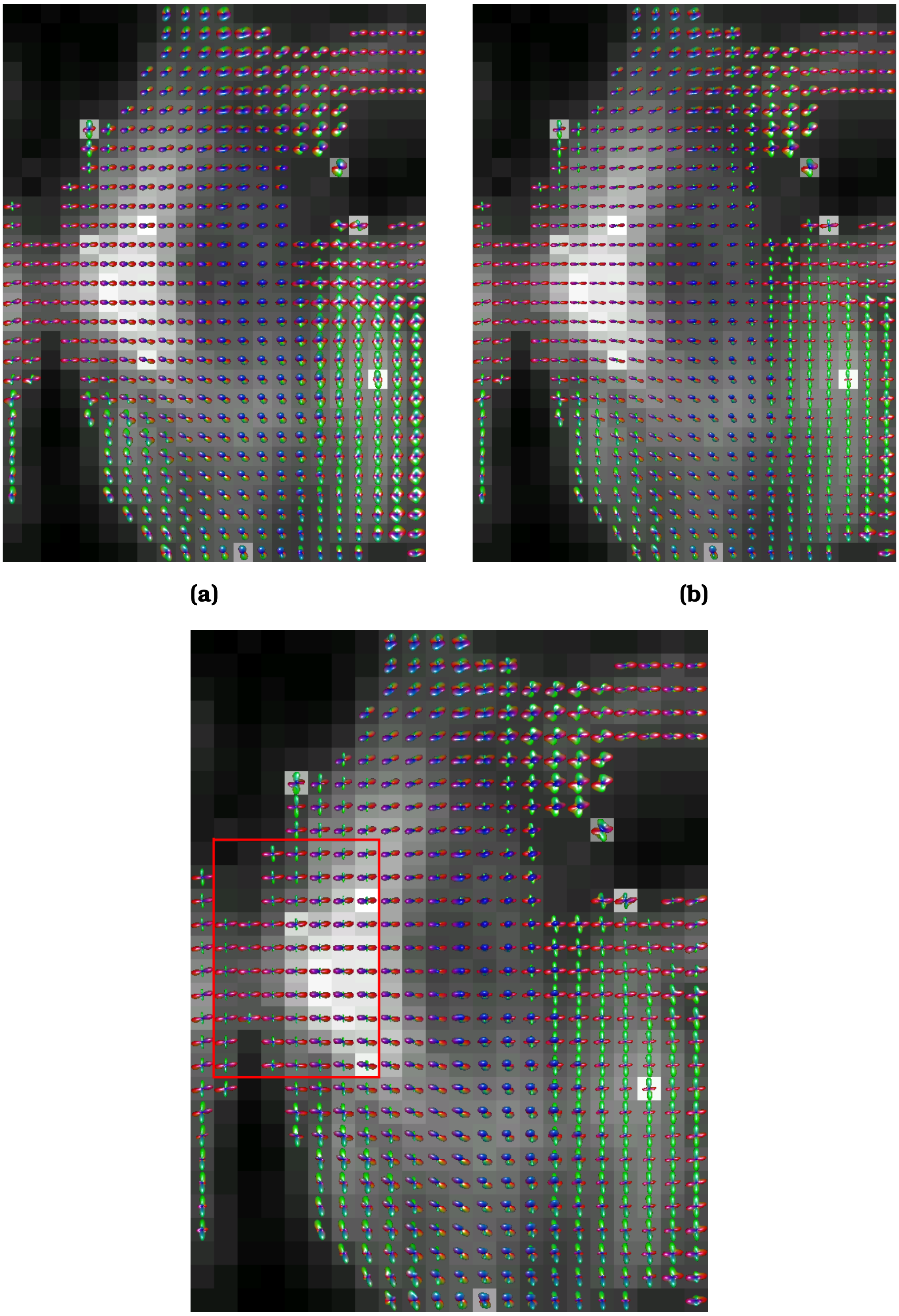
cpt: corticopontine tract

cst: corticospinal tract

slf: superior longitudinal fasciculus

str: superior thalamic radiation

Figure 8.6: CSIRO1: axial slice in the upper brain (slice 32). Color–coding representation plus FA map. The ROI is marked in red together with some tracts of interest (yellow).



(a)

(b)

(c)

Figure 8.7: CSIRO1: detail of the slected ROI in axial slice 32, for (a) Q–Balls, (b) the DOT, and (c) the cOPDT. The third picture has been enlarged for illustration purposes. Glyphs are colored following color–coding conventions for each axis, and the LMMSE–15 has been used to pre–process the DWI.

color–coding allows a very intuitive representation of the structures being imaged, regardless on the poor resolution (at least, compared to BWH1 or BWH2) of this volume. Fig. 8.7 shows glyphs for the selected ROI. Only Q–Balls, the DOT, and the cOPDT are analyzed, since the OPDT and the pOPDT show virtually identical

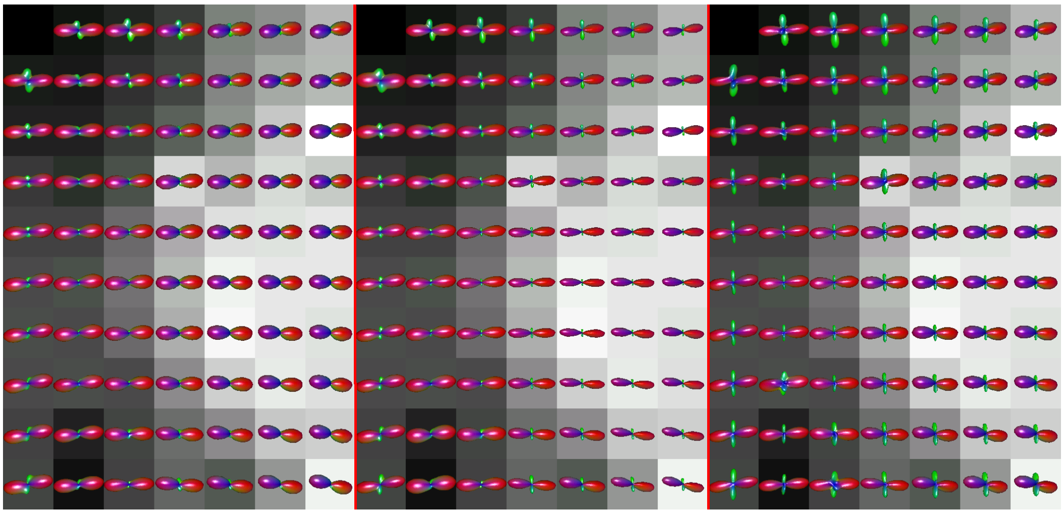


Figure 8.8: CSIRO1: Comparison of estimators in a small ROI inside axial slice 32. From left to right: Q–Balls, the DOT, and the cOPDT.

results to the cOPDT in this case. As can be seen, all the estimators are able to correctly resolve the structures coexisting in each region in all cases. Consider the third picture: on the left–hand side, the cingulum and the corpus callosum are distinguished, even when the corpus callosum precludes this other structure in the color–coding in Fig. 8.6. In the center of the ROI, the corticopontine tract (blue) is correctly discriminated from the corpus callosum (red). In the right–hand side, the superior longitudinal fasciculus is also separated from the transverse fibers (red) entering the ROI from its right border.

Fig. 8.8 shows a zoom of the region highlighted in the third picture of Fig. 8.7. The crossing between the cingulum and the corpus callosum is perfectly visible in all voxels of the picture corresponding to the cOPDT. For the DOT, such crossing is only resolved for some voxels, but not for all of them. Yet, when the crossing is detected, the direction corresponding to the cingulum (green) is difficult to appreciate when compared to the corpus callosum (red). For Q–Balls, the cingulum is resolved only for isolated voxels, mainly at the top of the corresponding picture. This result stresses the advantages of OPDT–related estimators when compared to

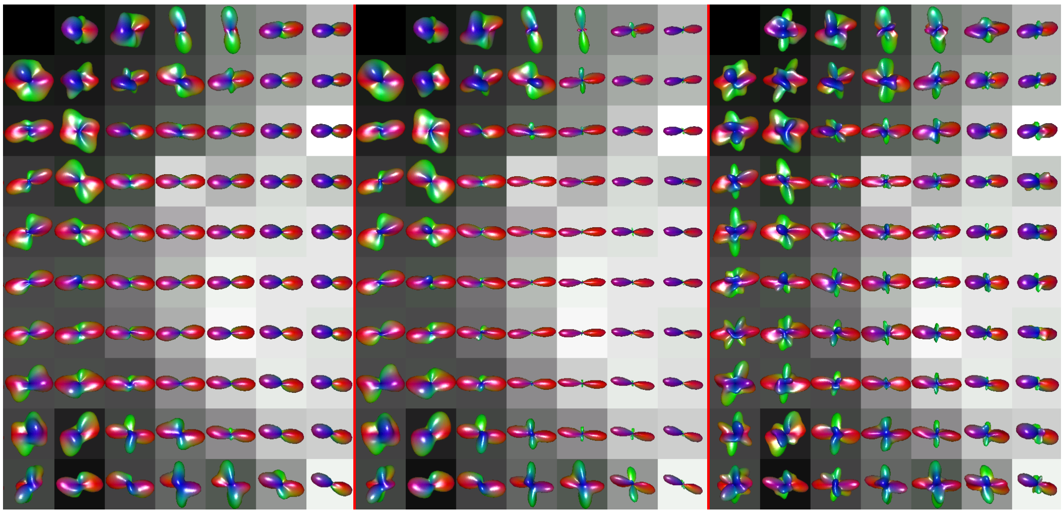
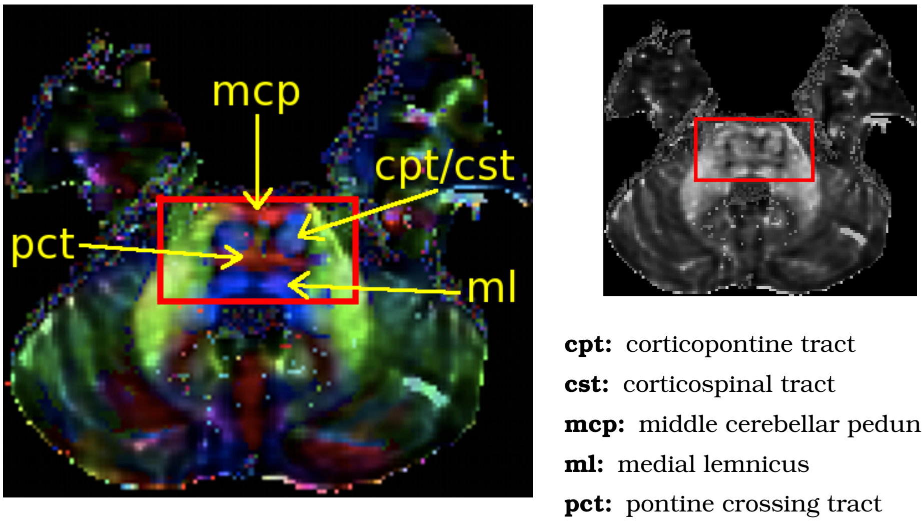


Figure 8.9: CSIRO1: Comparison of estimators in a small ROI inside axial slice 32, without LMMSE filtering. From left to right: Q–Balls, the DOT, and the cOPDT.



cpt: corticopontine tract

cst: corticospinal tract

mcp: middle cerebellar peduncle

ml: medial lemnicus

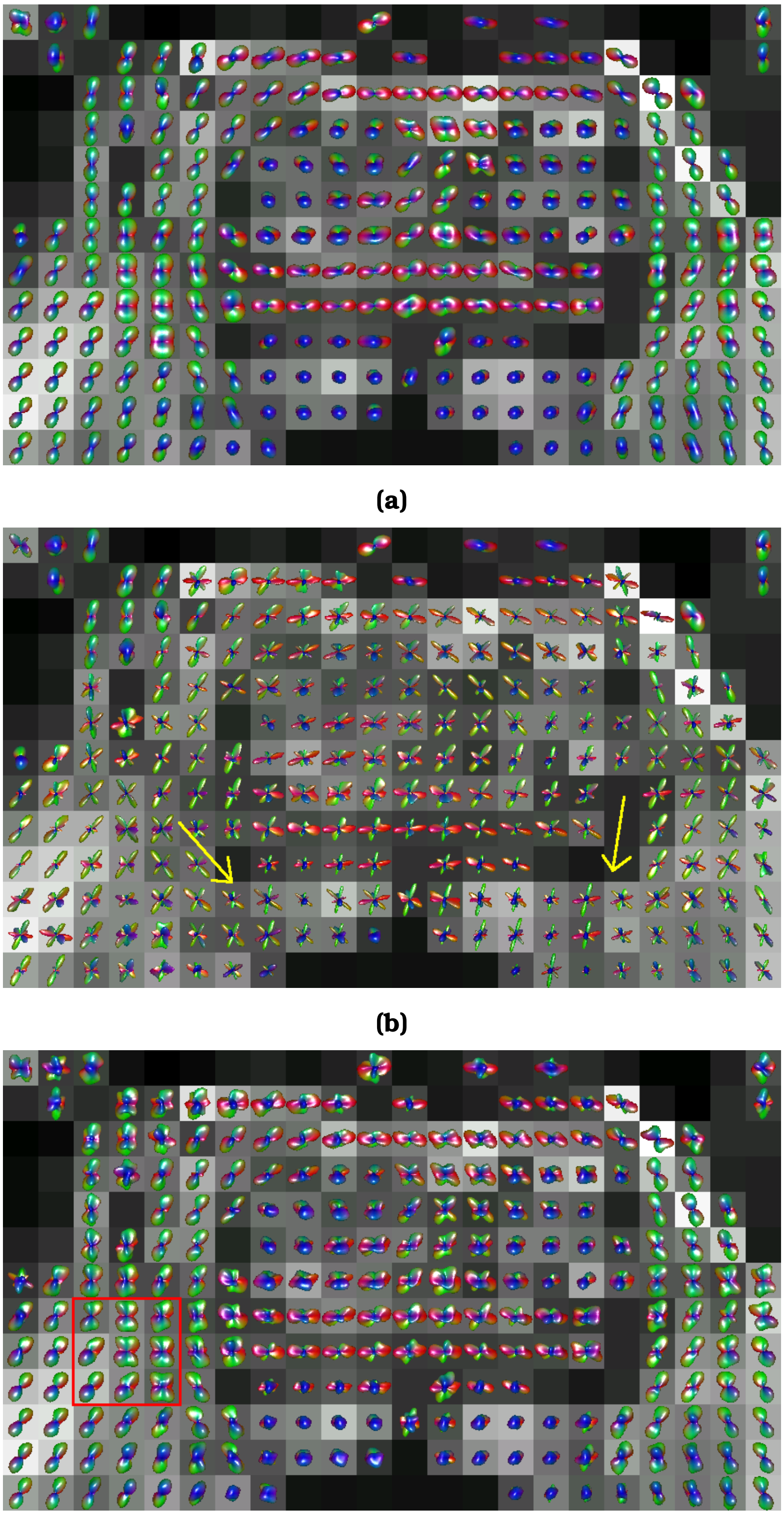
pct: pontine crossing tract

Figure 8.10: BWH1: axial slice in the bottom brain (slice 15). Color–coding representation plus FA map. The ROI is marked in red together with some tracts of interest (yellow).

the DOT or Q–Balls, even for large *b*–values. As an additional experiment, Fig. 8.9 shows complementary results to those of Fig. 8.8 when the LMMSE pre–processing is suppressed. Once again, the results are much less smooth and difficult to interpret without the denoising process. However, Q–Balls is still able to yield a smooth probability field, and only isolated voxels suffer from the effect of noise. With the DOT, now very few voxels account for both the cingulum and the corpus callosum, and the crossings have mostly disappeared. With the cOPDT, these crossings are well preserved in a high percentage of voxels inside the window: despite the distortion in the lobes of the orientation functions, the underlying information on the neural architecture is still intelligible. Two conclusions may be inferred: first, the cOPDT is able to discern complex architectures even in the presence of large amounts of noise (this is consistent with the results reported in previous Chapters). Even so, the second conclusion is the great advantage of filtering the DWI before estimating fiber populations. The LMMSE–*N* technique not only preserves the structural information, but it indeed enhances it, revealing information which remains occult in the original data sets.

8.3 Volume BWH1

The next set of experiments is carried out over data set BWH1. It may be considered also a HARDI volume, since it was acquired with a large number of sensitizing gradient directions (51). However, its *b*–value (700 s/mm^2) is quite smaller than the numbers typically used in HARDI estimation, which translates in a very poor angular contrast compared to CSIRO1. An axial slice of this volume, corresponding to the bottom of the brain, is represented in Fig. 8.10. The spatial resolution of this DWI data set is higher than that of CSIRO1, yielding smoother representations of the fiber bundles of interest by means of both color coding an FA maps. The main drawback of this volume, together with its reduced *b*–value, is the presence of a number of outliers in its whole extent. They can be recognized as isolated voxels with non–uniform values in the previous figure, and most of them persist even after the volume has been filtered. The ROI indicated in Fig. 8.10 has been analyzed in Fig. 8.11 with Q–Balls, the DOT, and the cOPDT (at first instance). Compared to the results previously reported in this Chapter, now it is the DOT



(a)

(b)

(c)

Figure 8.11: BWH1: detail of the slected ROI in axial slice 15, for (a) Q–Balls, (b) the DOT, and (c) the cOPDT. The slice has been decimated by a factor 2 for the sake of clarity. Glyphs are colored following color–coding conventions, and the LMMSE–15 has been used to pre–process the DWI.

which presents a great number of spurious lobes. Moreover, the representation it provides is very difficult to interpret, and yet it does not seem to correspond in

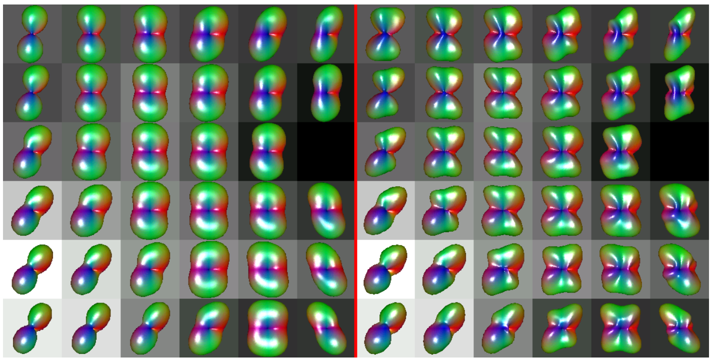


Figure 8.12: BWH1: Comparison of estimators in a small ROI inside axial slice 15. Q–Balls (left), and the cOPDT (right) are compared.

a trivial way to the anatomies described in Fig. 8.10. In particular, the corticopontine tract/corticospinal tract and the medial lemnicus are highly distorted, so that it is very difficult to identify the blue structures (transverse to the plain of the image) representing them. Yet, at the bottom of the image (see the voxels indicated by arrows), the orientation information is completely unintelligible, due to the existence of outlying lobes. Regarding Q–Balls, it is able to yield the smoothest representation, with regular glyphs matching the expected orientations. Notwithstanding, this estimator is not able to correctly account for the great number of fiber crossings present in this ROI due to the complex anatomy in this region. These results are consistent with the discussions on previous Chapters over again: for low b–values, neither the DOT nor Q–Balls show an optimum behavior, and it is precisely in these scenarios where OPDT–based techniques are more advantageous. Consider the region marked in the third picture of Fig. 8.11, a detail of which is represented in Fig. 8.12. Only Q–Balls and the cOPDT are compared, since the DOT does not seem adequate in this scenario. As it can be seen, Q–Balls is not able to account for the crossings the cOPDT evidences. Moreover, Q–Balls yield estimates suggesting that the fiber bundles run vertically from the top of the image to its bottom part. This way, the pattern formed by the middle cerebellar peduncle and the pontine crossing tract is mostly lost with Q–Balls, while the cOPDT is able to correctly resolve these structures for practically all voxels.

At this point, it is useful to compare the OPDT–based estimators, since their behavior is no longer the same for BWH1. Fig. 8.13 shows that the performance of both the OPDT and the cPDT is virtually the same. Comparing the glyphs at each voxel, it is impossible to distinguish between the results provided by these two estimators. For the pOPDT, noticeable differences may be appreciated. The orientation information for this estimator is more distorted, so that the positions

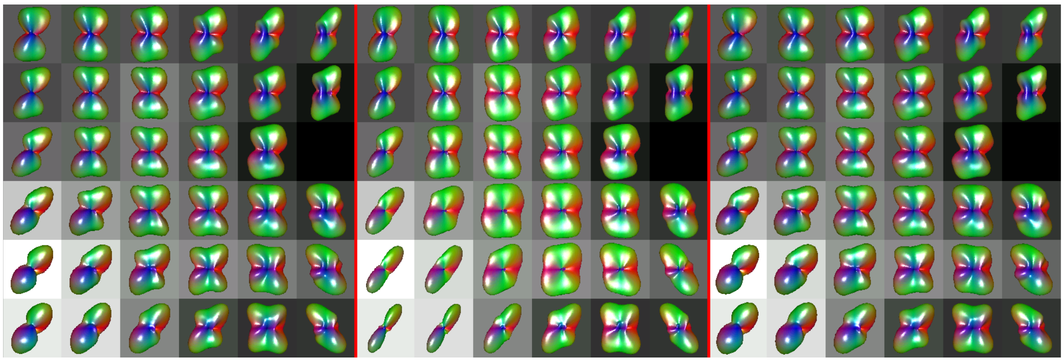


Figure 8.13: BWH1: Comparison of estimators in a small ROI inside axial slice 15. The OPDT (left), the pOPDT (center), and the cOPDT (right) are compared.

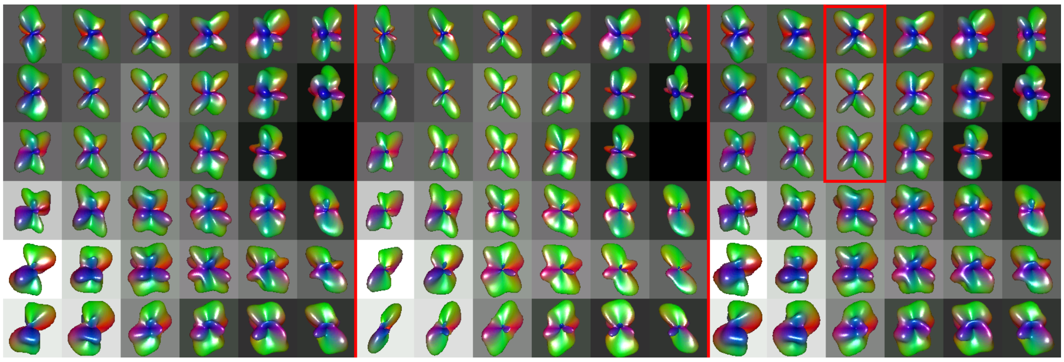
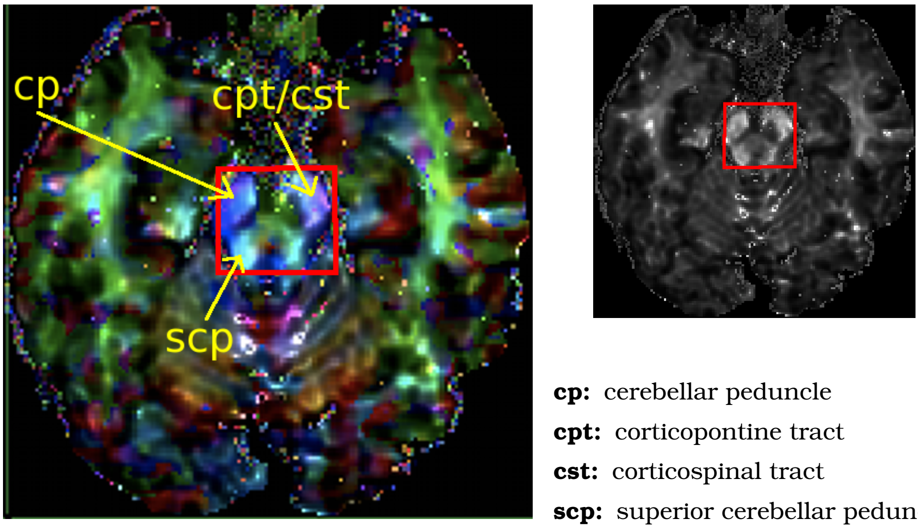


Figure 8.14: BWH1: Comparison of estimators in a small ROI inside axial slice 15, without LMMSE filtering. The OPDT (left), the pOPDT (center), and the cOPDT (right) are compared.

of the maxima (wherever two maxima can be identified) result more ambiguous. Besides, the spatial coherence of the glyphs in this case is not so obvious. The previous results correlate well with the experiments performed in Chapters 6 and 7 with synthetic data: for low b–values (1000 s/mm^2 or less), the performance of Q–Balls and the DOT is very poor, and only OPDT–based estimators are able to resolve complex architectures even for small crossing angles. Comparing between them, the cOPDT does not show any particular advantage over the OPDT itself, although the pOPDT shows a slightly worse behavior. From the previous Chapters, a factor which can differentiate the performance of these techniques is the amount of noise. Hence, the experiment in Fig. 8.13 is repeated without the LMMSE–15 pre–processing so that the SNR is drastically reduced, and the results presented in Fig. 8.14. Most of the voxels present unintelligible orientations, and the structural information is mostly lost. Nevertheless, for some isolated voxels, like those highlighted in the right–most picture, the lobes corresponding to directions of interest are better defined than in the filtered volume (see Fig. 8.13). As opposed to the previous discussion, this is the first adverse effect observed in the LMMSE technique: on the average, LMMSE–N is still able to improve the estimation and reveal hidden anatomies, but for some voxels it may as well blur the orientation information. Yet, comparing Figs. 8.13 and 8.14, it remains evident that the denoising stage is essential to achieve accurate estimates of fiber populations. Therefore, in



cp: cerebellar peduncle

cpt: corticopontine tract

cst: corticospinal tract

scp: superior cerebellar peduncle

Figure 8.15: BWH1: axial slice in the bottom brain (slice 23). Color–coding representation plus FA map. The ROI is marked in red together with some tracts of interest (yellow).

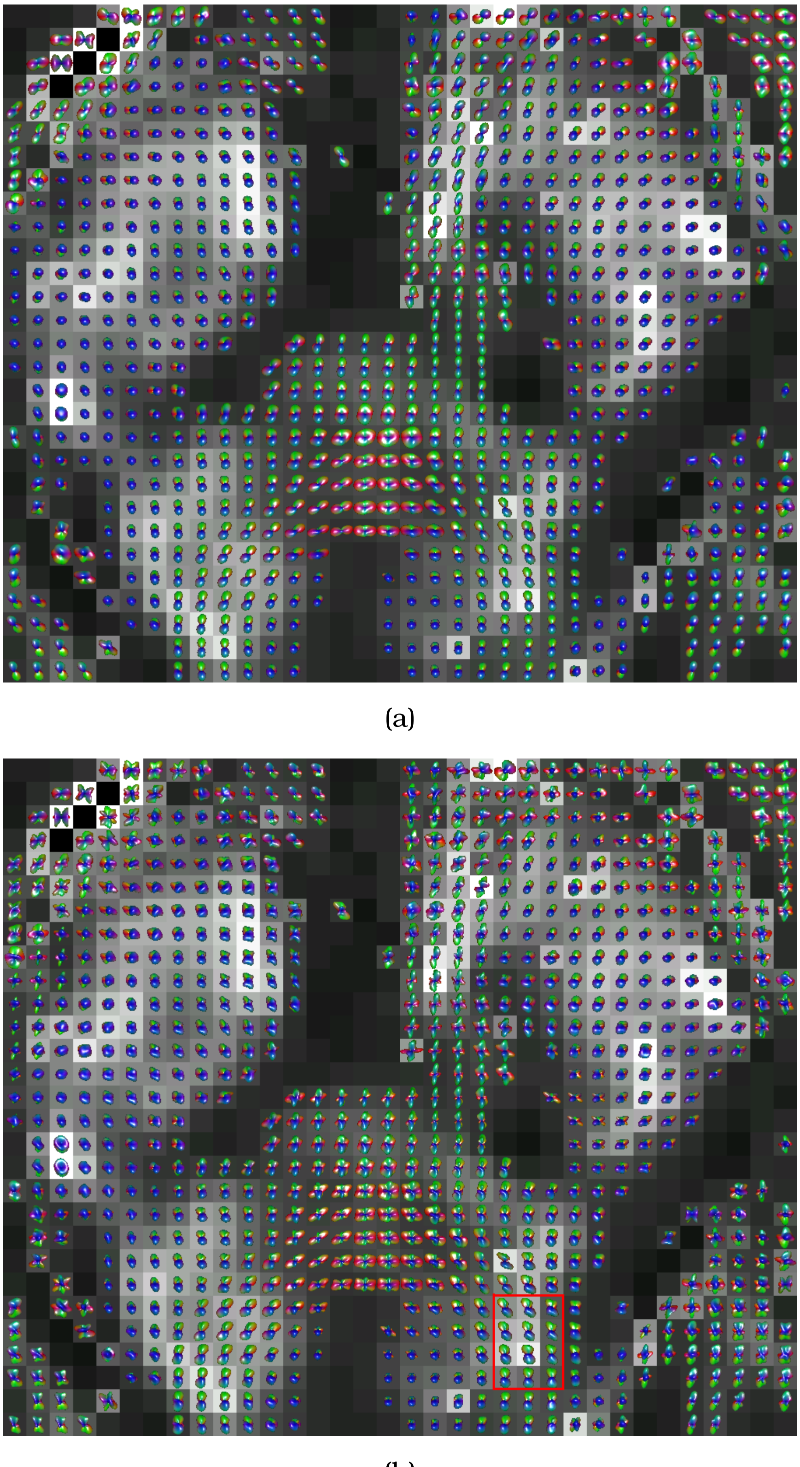


Figure 8.16: BWH1: detail of the slected ROI in axial slice 23, for Q–Balls (a), and the cOPDT (b). Glyphs are colored following color–coding conventions, and the LMMSE–15 has been used to pre–process the DWI.

what follows only results from filtered data are presented, except where noticed.

In the next experiment, the axial slice represented in Fig. 8.15 is analyzed. The aforementioned outliers and artifacts are especially noticeable in this slice, although the region studied is not very damaged (at least in the voxels com-

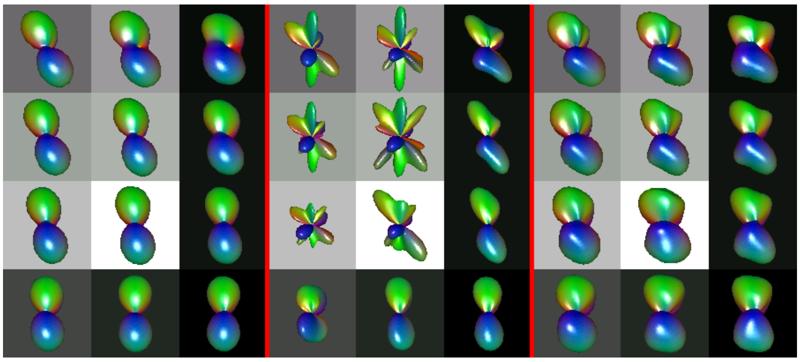
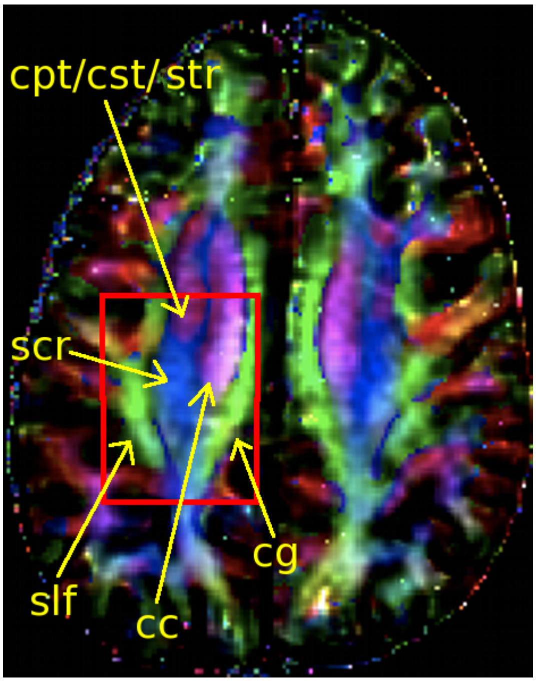
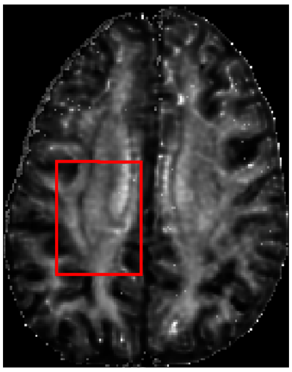


Figure 8.17: BWH1: Comparison of estimators in a small ROI inside axial slice 23. Q–Balls (left), the DOT (center), and the cOPDT (right) are compared.

prising tracts of interest). The corresponding glyph representation is given in Fig. 8.16 for Q–Balls and the cOPDT (for the DOT, the results are similar to those in Fig. 8.11 (b), and so they are omitted). The first comment is that, over again, the cOPDT (or, in general, HARDI analysis) reveals neural architectures which are not adequately described by DTI techniques: in Fig. 8.15 , the color–coded image suggests that the tracts of interest are mainly aligned wit the 'z' axis, in the transverse direction to the plane of the image. On the contrary, Fig. 8.16 shows that in fact the underlying architecture is more complex, especially in the bottom half. The decussation of the superior cerebellar peduncle, which is hardly visible in Fig. 8.15 as a small red spot, is perfectly visible in Fig. 8.16, and the crossings with the transverse (blue) structures are well resolved. Additionally, Fig. 8.17 shows a detailed view of the small ROI marked in Fig. 8.16 (b). The DOT is included now for comparison purposes. However, this estimator is again inadequate: in many voxels, important lobes appear in the orientation information function that prevent its correct interpretation. Only the glyphs on the top–right corner have been





cc: corpus callosum

cg: cingulum

cpt: corticopontine tract

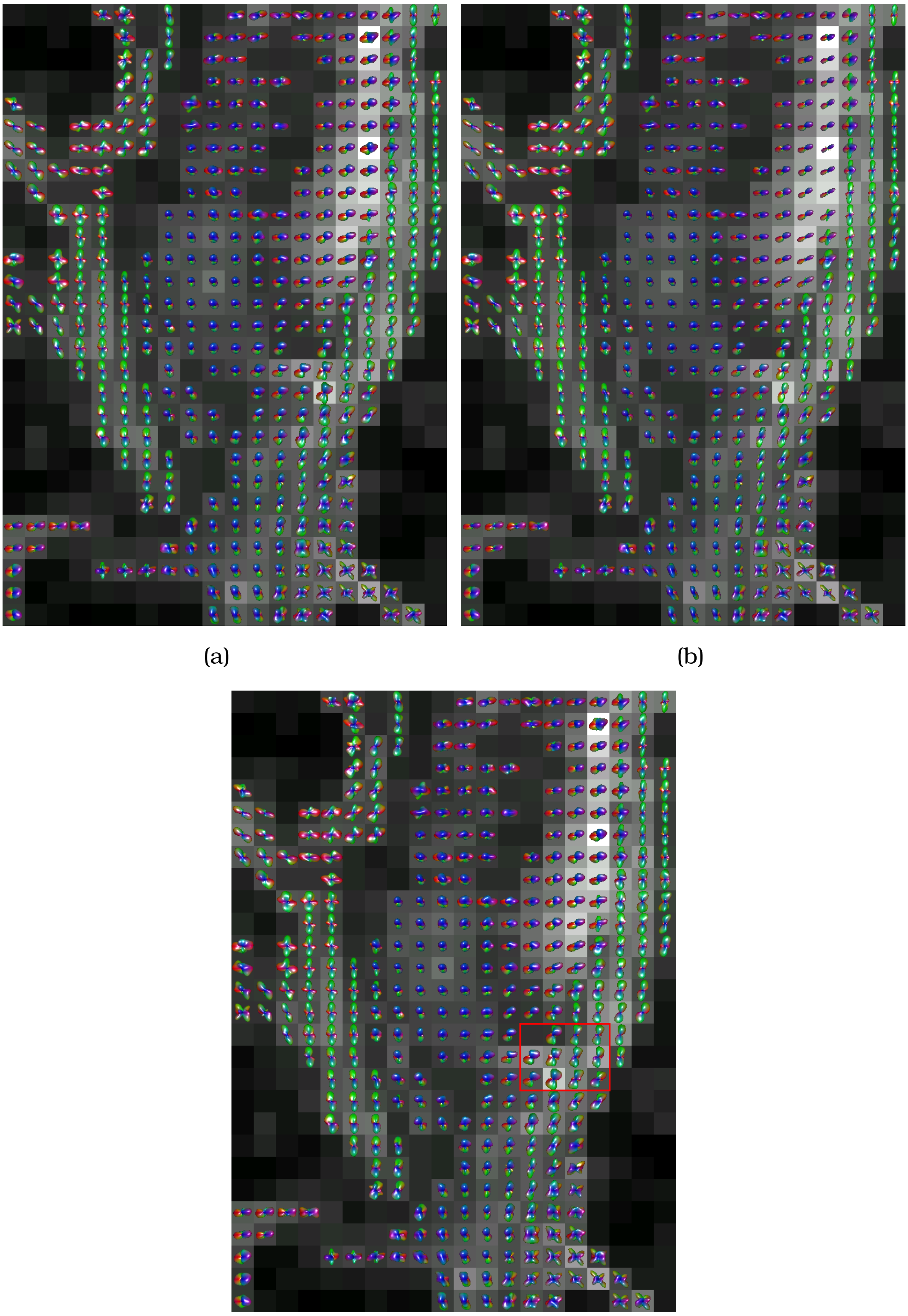
cst: corticospinal tract

slf: superior longitudinal fasciculus

scr: superior corona radiata

str: superior thalamic radiation

Figure 8.18: BWH1: axial slice in the upper brain (slice 52). Color–coding representation plus FA map. The ROI is marked in red together with some tracts of interest (yellow).



(a)

(b)

(c)

Figure 8.19: BWH1: detail of the slected ROI in axial slice 52, for the OPDT (a), the pOPDT (b), and the cOPDT (c). The slice has been decimated by a factor 2 for the sake of clarity. Glyphs are colored following color–coding conventions, and the LMMSE–15 has been used to pre–process the DWI.

correctly estimated. With regard to Q–Balls, the glyphs are highly blurred, so that the fiber crossings are not properly described. The cOPDT, in turn, is able to dis-

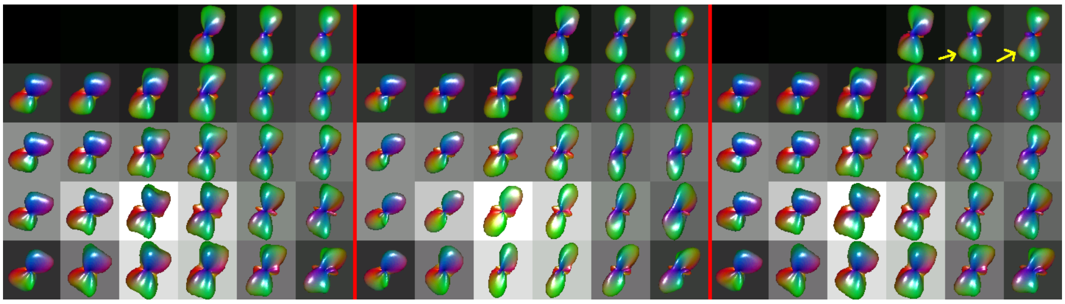


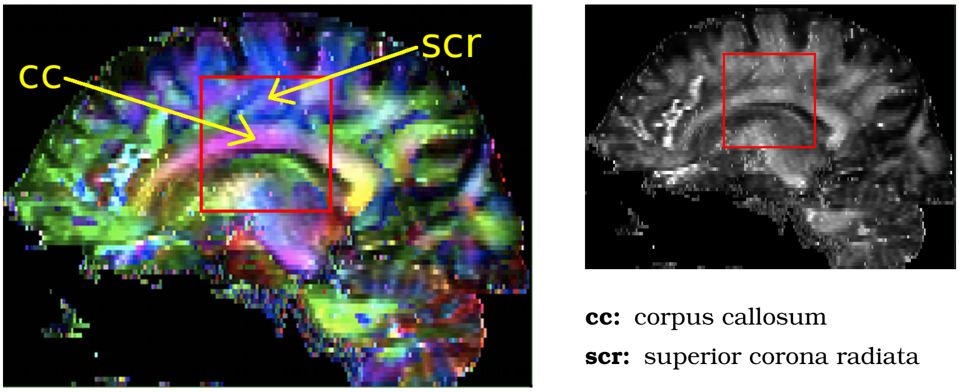
Figure 8.20: BWH1: Comparison of estimators in a small ROI inside axial slice 52. The OPDT (left), the pOPDT (center), and the cOPDT (right) are compared.

tinguish two maxima in the orientation function for most of the voxels analyzed, corresponding to the pattern formed by the superior cerebellar peduncle and its decussation. This experiment, together with those performed for slice 15, evidences the higher performance of OPDT–related techniques to describe complex architectures for low *b*–values. In what follows, the discussion is mostly centered on the comparison of these techniques, i.e. the OPDT, the cOPDT, and the pOPDT.

To begin with, the axial slice depicted in Fig. 8.18 is considered. This slice is analogous to that in Fig. 8.6 for volume CSIRO1, and represents the imbricated anatomies in the upper brain, where the corpus callosum and the cingulum cross in transverse directions. Besides, the superior corona radiata, the superior longitudinal fasciculus, and other important tracts overlap in the ROI studied. Compared to Fig. 8.6, now the axial resolution of the volume is much higher, so both the color–coding and the FA map provide rich information on these structures. As always, a glyph representation of the selected ROI is given in Fig. 8.19, where results obtained with the OPDT, the pOPDT, and the cOPDT are represented. As it can be seen, the three estimators yield very similar outcomes. Although not so evident as in Fig. 8.7, the crossings between the cingulum and the corpus callosum are still visible, as well as the intersections of the superior longitudinal fasciculus with the transverse (red) structures entering the ROI on its left–hand side. The glyphs are able to account as well for the crossings between the corona radiata and the corpus callosum (on the top–right corner) or the corticopontine tract (on the top–center). In any case, this representation allows a more intuitive interpretation of the anatomies than DTI does.

The window highlighted in Fig. 8.19 (c) is expanded in Fig. 8.20 to allow a better comparison of the estimations. The results are very similar to those presented in Fig. 8.13, which in turn are consistent with the experiments performed over synthetic data sets: The OPDT and the cOPDT behave quite similar, while the pOPDT seems to produce a higher blurring of the orientation information. It translates in that the crossing between the cingulum and the corpus callosum, which is perfectly visible with the OPDT and the cOPDT, is missing with this other estimator. Moreover, this lost information reduces the spatial coherence of the results with the pOPDT, yielding visually less adequate results. Comparing now the OPDT and the cOPDT, their behavior is completely comparable. Only very subtle differences can be detected in some glyphs, which are marked in the third picture with yellow arrows. The cOPDT seems to better resolve the structures for these voxels, although the improvement is not very significant.

For the next experiment, a sagittal slice is chosen instead of the axial view selected for all the previous experiments. It is represented in Fig. 8.21. The color–coding in this figure shows that a number of tracts are present inside the ROI, crossing in all possible directions. The corresponding glyphs are shown in Fig. 8.22 for the cOPDT. As can be seen, the glyphs corresponding to the corpus callosum



cc: corpus callosum

scr: superior corona radiata

Figure 8.21: BWH1: sagittal slice in the mid brain (slice 147). Color–coding representation plus FA map. The ROI is marked in red together with some tracts of interest (yellow).

are not completely transverse to the plain of the image, but their direction is oblique. This behavior accurately fits the color–coding in Fig. 8.21, where the corpus callosum is not represented in pure red but in magenta instead. Besides, the analysis by means of the cOPDT reveals the presence of the cingulum, which crosses the image from left to right (and therefore is represented in green) over the corpus callosum. Moreover, the crossings of this structure with the superior corona radiata are accounted for in the majority of voxels represented. This feature is especially relevant at the sight of Fig. 8.21, for which the cingulum is not visible at all. The corresponding detailed view for the window depicted in Fig. 8.22 is presented in Fig. 8.23 for the three estimators. In this case, the differences be-

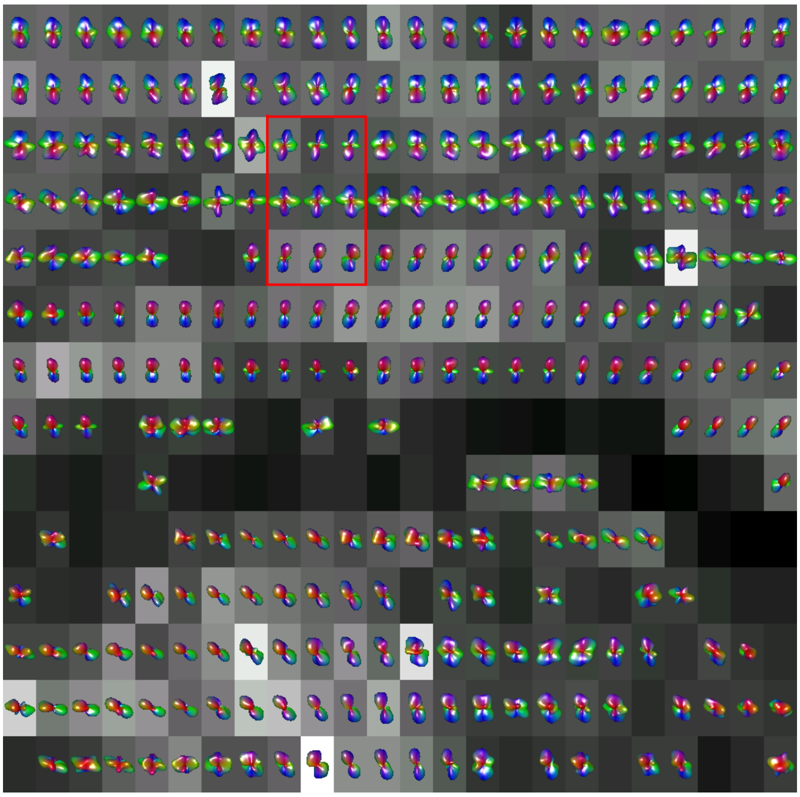


Figure 8.22: BWH1: detail of the slected ROI in sagittal slice 147, for the cOPDT. The slice has been decimated by a factor 2 for the sake of clarity. Glyphs are colored following color–coding conventions, and the LMMSE–15 has been used to pre–process the DWI.

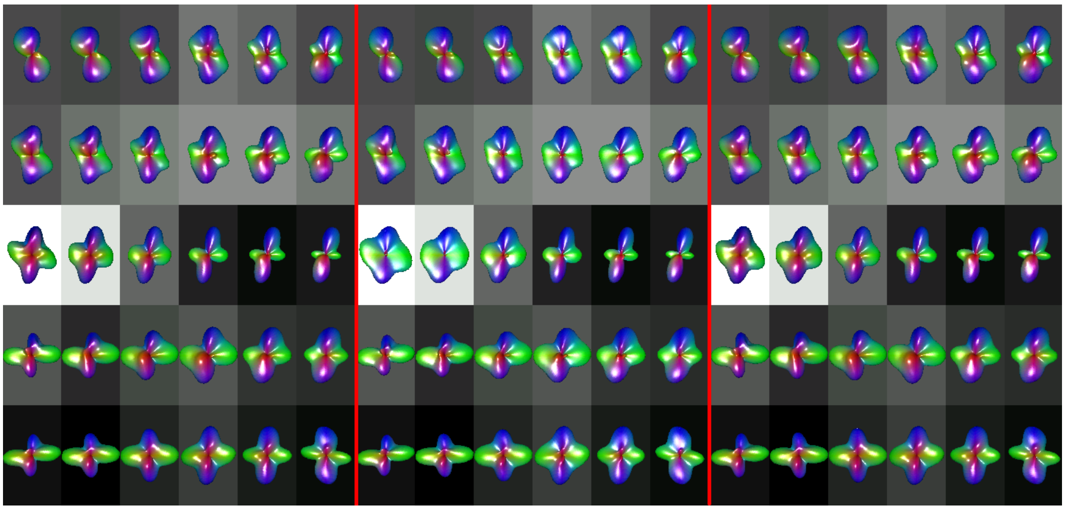
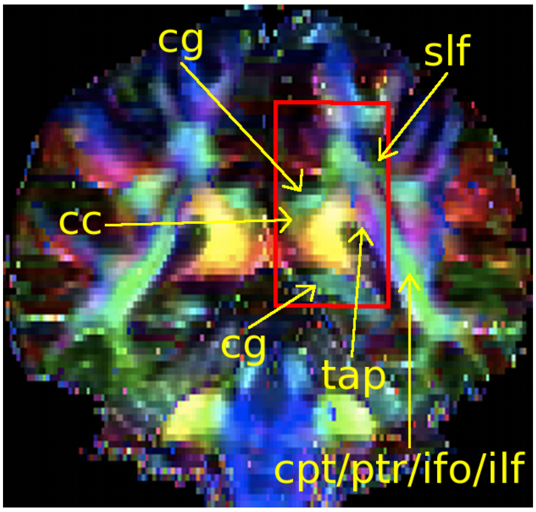
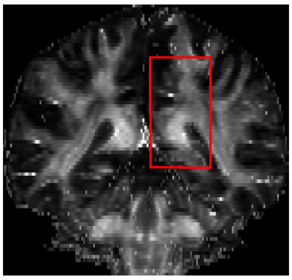


Figure 8.23: BWH1: Comparison of estimators in a small ROI inside sagittal slice 147. The OPDT (left), the pOPDT (center), and the cOPDT (right) are compared.

tween the OPDT, the pOPDT, and the cOPDT are negligible. All of them are able to resolve the crossings between the corona radiata and the cingulum, although the lobes of the orientation information functions are quite blurred in the top of the image. In this case, fiber bundles cross in large angles (nearly 90^o), so it was expected *a priori* that their behaviors have to be approximately the same. Despite of the blurring in the OPDF, the HARDI analysis enables to distinguish the cigulum, which was completely hindered in the DTI study.

In the final example with data set BWH1, the coronal slice in Fig. 8.24 is analyzed. This ROI comprises a number of tracts of interest rendering a very complex architecture. The corpus callosum is represented in yellow since it is crossing the plane of the image in an oblique angle. The cingulum, which follows a completely transverse direction along the ‘y’ axis, appears as a well differentiated green spot





cc: corpus callosum

cg: cingulum

cpt: corticopontine tract

ifo: inferior fronto-occipital fasciculus

ilf: inferior longitudinal fasciculus

ptr: posterior thalamic radiation

slf: superior longitudinal fasciculus

tap: tapetum

Figure 8.24: BWH1: coronal slice in the mid brain (slice 88). Color–coding representation plus FA map. The ROI is marked in red together with some tracts of interest (yellow).

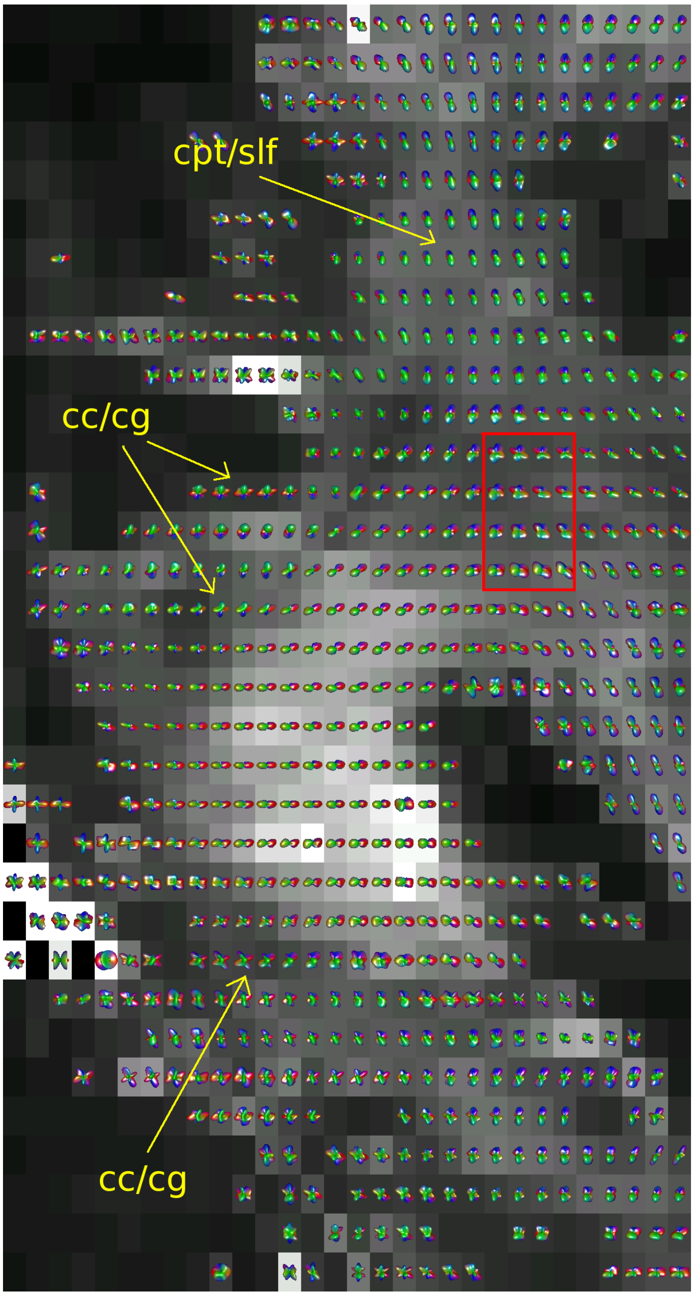


Figure 8.25: BWH1: detail of the slected ROI in coronal slice 88, for the cOPDT. Glyphs are colored following color–coding conventions, and the LMMSE–15 has been used to pre–process the DWI.

next to the corpus callosum. The other tracts in the image, such as the tapetum, the superior longitudinal fasciculus, or the corticopontine tract, have very diverse configurations.

The glyphs obtained with the cOPDT for the corresponding ROI are presented in Fig. 8.25. The corpus callosum is accurately resolved, with glyphs pointing in an oblique direction as expected. The cingulum is also well accounted, including its intersection with the corpus callosum, which is represented by the cross–shaped glyphs highlighted in the image. Although their visibility is reduced because the corticopontine tract (and the other associated tracts) is transverse to the plane of the image, the intersections of this structure with the longitudinal fasciculus is also visible in the majority of the voxels. Voxels corresponding to the intersection

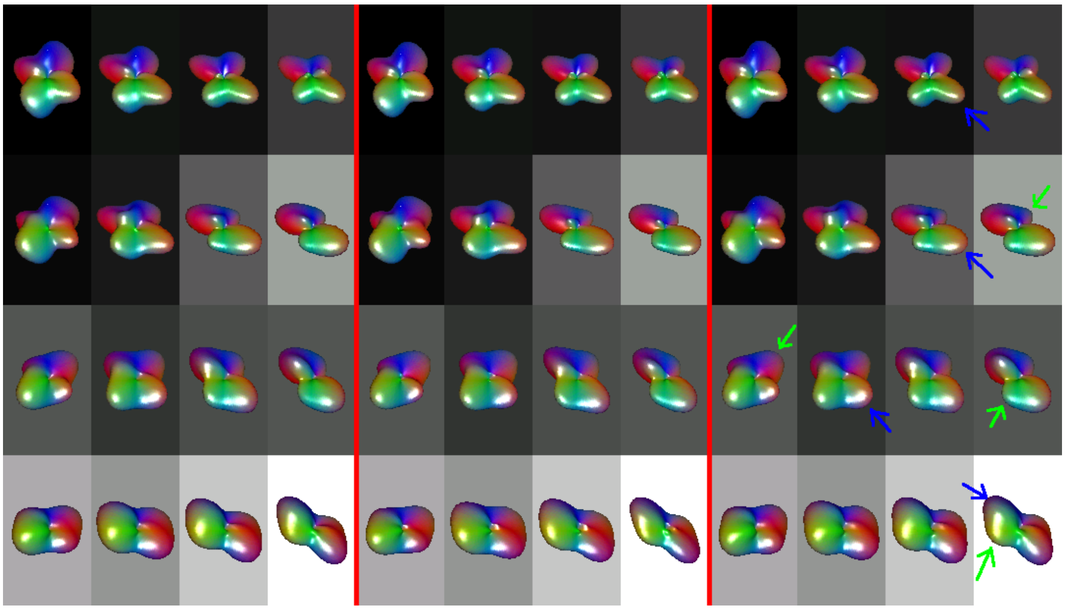


Figure 8.26: BWH1: Comparison of estimators in a small ROI inside coronal slice 88. The OPDT (left), the pOPDT (center), and the cOPDT (right) are compared.

of the tapetum with the corpus callosum have been detailed inside the ROI depicted. Fig. 8.26 shows a zoom of this region. As suggested by the color–coding image, the direction of the tapetum is not completely vertical, but instead it has a component in the ‘x’ (red) direction. Such direction is represented in Fig. 8.26 by the maxima marked with a blue arrow. The crossings of this anatomy with the corpus callosum (whose direction is marked with green arrows) is resolved by all the estimators. In fact, the behavior of all of them is quite similar. Much like in the previous experiment, the angle formed by the corpus callosum and the tapetum is large enough to be detected by all the OPDT–like techniques.

8.4 Volume BWH3

The last experiment in this Chapter is to be carried out over data set BWH3. As stated in Chapter 1, this volume brings on an additional difficulty, since the raw data is available only as a MatLab `.mat` file. Moreover, only the attenuation signal was provided, so no information regarding the T2 baseline image can be used. Although theoretically this should not be a problem for the estimation of fiber populations (which is based only on the attenuation signal), there is an important side effect related to this issue: the LMMSE–N technique described in Chapter 5 requires to use the baseline to account for the diffusion model, so it does not apply in the current scenario. Yet, when the gradient images are divided by the baseline, the statistics of noise are obviously modified. It was argued in Chapter 6 that the SNR in the baseline is far larger than in the gradient images, so the noise can still be considered as Rician. Even so, the baseline shows very different values at each image location, and as a consequence the noise power in the attenuation signal is highly variable (it is divided by the squared baseline at each voxel), so the filtering techniques presented are not adequate. As a consequence, raw (noisy) data has to be managed with BWH3. An axial slice in the bottom brain, corresponding to the same anatomy as that analyzed in Fig. 8.10, is studied in Fig.8.27 for data set BWH3, with the cOPDT. The results are much less smooth than those reported for BWH1, since no filtering can be performed. Although the directions of tracts

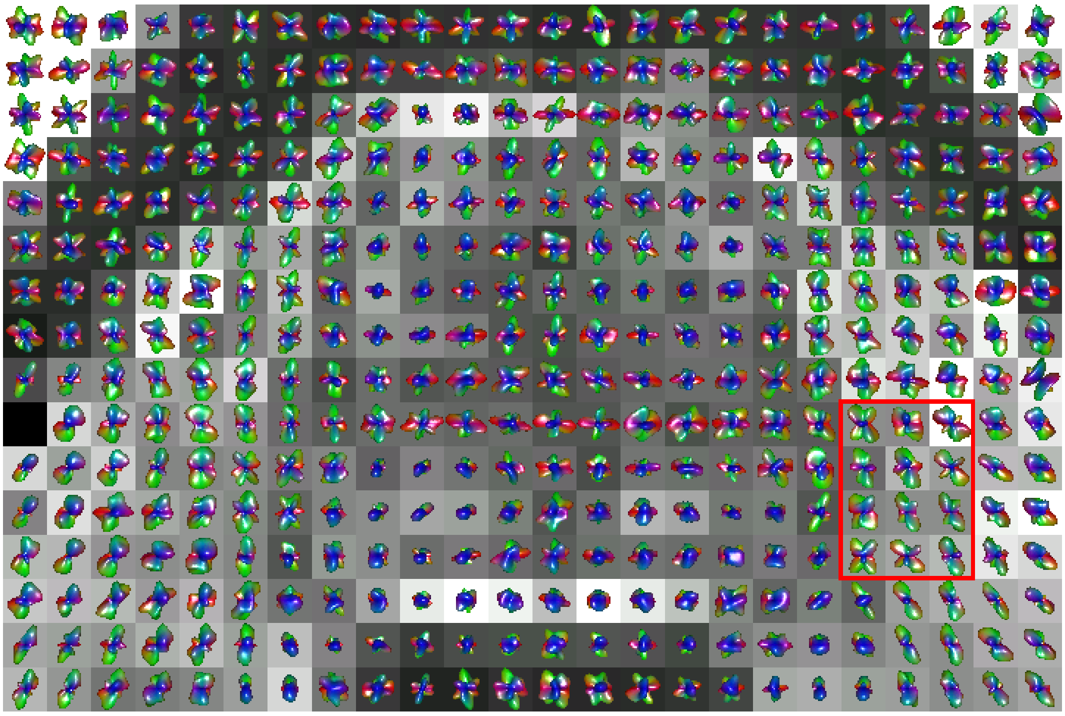


Figure 8.27: BWH3: detail of a ROI in axial slice 19, corresponding to the same anatomy depicted in Fig. 8.10, for the cOPDT. Glyphs are colored following color–coding conventions. The LMMSE–15 technique cannot be applied in this case.

of interest are in general terms well resolved, the interpretation of the results is more difficult than it was for BWH1 (see Fig. 8.11), even when the *b*–value of BWH3 is larger (1000 s/mm^2). Fig. 8.28 shows a zoom of the window represented in Fig. 8.27. In this case, it is interesting to include Q–Balls and the DOT in the comparison. As can be seen, Q–Balls is not able to resolve the fiber crossings in practically any voxel (unless for those in the left column, which are especially easy to detect). On the contrary, the DOT seems to be useful in this particular situation. It is able to represent all the fiber crossings the other estimators distinguish, with very few outlying lobes. Nonetheless, the directions estimated for the crossing fibers are not always completely coherent with those estimated by the other techniques. As an example, consider the direction marked on one of the glyphs with a red arrow. By simple visual inspection, it seems clear that the detected fiber is not following the same orientation provided by the other approaches. From the experiments with synthetic data, it is expected that the error committed by the DOT is larger than that of OPDT–based estimators; besides, in the window chosen for this analysis, the behavior of the DOT is especially accurate, but this is not the case for the rest of the ROI. Comparing now the OPDT, the cOPDT, and the pOPDT, their behavior is over again quite similar, although some differences can now be identified. The most remarkable one is in the glyph pointed out with a green arrow. For this specific position, both the DOT and the OPDT show outlying lobes which preclude the directions of fibers. With regard to the pOPDT, it is unable to identify the crossings, so that it reveals one unique fiber bundle much like as Q–Balls. On the contrary, the cOPDT reveals three crossing fibers, whose orientations are completely coherent with their surroundings. Hence, it seems reasonable to assume that these orientations correspond to actual anatomies. Moreover, the glyphs neighboring that marked with the green arrow unveil this same structure, which is only visible with the OPDT and the cOPDT, but not with the pOPDT (or with Q–Balls or the DOT either).

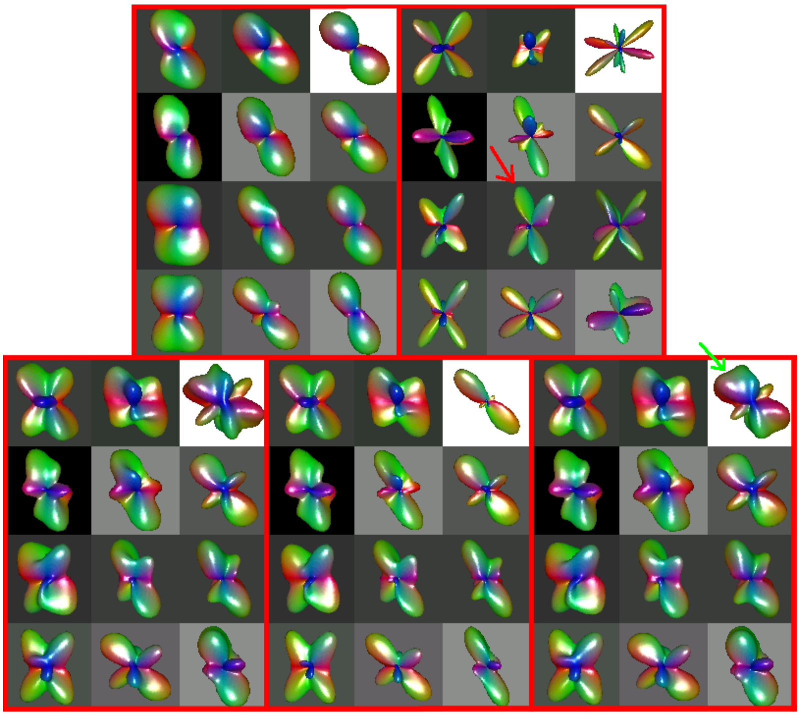


Figure 8.28: BWH3: Comparison of estimators in a small ROI inside axial slice 19. Q–Balls (top–left), the DOT (top–right), the OPDT (bottom–left), the pOPDT (bottom–center), and the cOPDT (bottom–right) are compared.

8.5 Discussion

This Chapter has allowed to validate over real data most of the premises which motivated the present research work. Perhaps, the most important conclusion from the experiments carried out is that in fact HARDI techniques provide highly valuable information beyond DTI, revealing neural architectures which remain occult with this other approach. This comment holds for practically all the results presented. At the same time, it has been shown that the brain regions for which the description of complex architectures is useful are numerous. The fiber bundles involved in the study, such as the corpus callosum, the cingulum, the superior longitudinal fasciculus, the cerebellar peduncle, and others, are very relevant tracts with important functions. It has been evidenced that the study of these structures can benefit from HARDI estimation in a number of areas of interest, including the middle and the peripheral brain, axial views and sagittal/coronal slices.

The quantitative evaluation of the techniques presented has been carried out in previous Chapters, based on synthetic data on standard simplified scenarios. However, this analysis was centered on the accuracy for isolated voxels. In the present Chapter, the evaluation is merely qualitative due to the lack of a golden standard. On the other hand, the use of real data sets has allowed to test other important features such as the spatial coherence of the orientation information estimated and its anatomical correctness. Interestingly enough, the results obtained over real data are consistent with the conclusions derived from the study with synthetic data, which in part justifies the methodology for quantitative evaluation.

In particular, the advantage of using large b–values has become evident. Although it introduces a high attenuation, noticeably worsening the SNR, the improvement in the angular contrast compensate this drawback: in this case, all

HARDI estimators are able to yield appropriate results, providing very rich information which conventional DTI techniques cannot account for. The qualitative performance of all of them is quite similar, so from this analysis it would be difficult to assess the possible pros and cons of each one. Still, the quantitative validation carried out formerly advices the use of OPDT–based estimators, unless the architectures to resolve are well differentiated (i.e. fibers crossing or bending in large angles). Some illustrative experiments have been shown which suggest that in fact OPDT–like approaches may be advantageous even for large b–values, see Figs. 8.3 and 8.8. Of course, this assertion relates to the qualitative representation of estimations; for quantitative analyses, the advantage is even clearer.

For smaller b–values, the comparison clearly favors the techniques proposed in this dissertation. Although the orientation information is not so easy to interpret, due to the widening of the lobes of the OPDF, it is still intelligible. On the contrary, the DOT tends to render a large number of spurious lobes not always following true fiber bundles. As a consequence, the orientation information it provides is completely distorted. With Q–Balls, the shape of the ODF is still rather smooth, but it is not able to distinguish fiber crossings even for quite large angles. Besides, and what is more important, this approach may drive to ambiguities in the orientations of fibers under some circumstances (see Fig. 8.12). This result is consistent with the discussions following the numerical validation in previous Chapters.

With regard to the novel estimators presented in Chapter 7 (those estimating true probabilistic information, i.e. the OPDF), the OPDT and the cOPDT show a very similar behavior (once again, the quantitative comparison has to be used to argue which of them is better fitted for each particular situation). The pOPDT can yield poorer results in some cases (see Figs. 8.13 and 8.20), when fiber crossings in very small angles are present. However, in case the directions to resolve are well differentiated, their qualitative behavior is virtually the same.

Finally, it is worth stressing the adequacy or even the need for filtering DWI data sets prior to the estimation of fiber populations. It remains evident that this stage allows to obtain smoother results, and to improve their spatial coherence in all cases. Moreover, it may become essential to achieve interpretable estimates in most of cases. The exception for this rule is in Q–Balls, which is especially robust to noise. The denoising process may indeed carry out certain disadvantages, distorting the angular resolution at some isolated voxels (compare Figs. 8.13 and 8.14). Nonetheless, it has been shown that it is fundamental to attain appropriate results, especially for small b–values. Yet, the experiments performed widely report that the LMMSE–N technique is able to disclose neural architectures precluded by the effect of noise, so it is clearly advantageous in all situations.

9

Conclusions and future lines

A complete framework for the analysis, processing, and interpretation of diffusion MR images has been presented. It comprises the steps from the characterization of diffusion data in the MRI scanner to high level processing in order to infer fiber populations in the white matter. Chapter 2 was devoted to the study of the foundations of MRI and diffusion imaging. The review of the physical principles behind these techniques, mostly based on quantum mechanics, is far beyond the scope of the work presented. Instead, a comprehensive analysis focused on the mathematical and statistical characterization of this kind of data was performed. This study was centered not only in conventional acquisition techniques, but also on the analysis of the most recent approaches. Concretely, pMRI in diffusion imaging and HARDI have centered a major part of this dissertation. Although the use of these techniques is not widespread nowadays, their great potential and nice properties suggest that more attention will be paid to them in the near future.

Chapter 3 was focused on the statistical characterization of noise in diffusion (and conventional) MRI data sets. From the study carried out, the most important conclusion is the general division of MRI protocols into Rician–distributed and non–central Chi–distributed. In conventional MRI and most of the current pMRI protocols (SENSE, SMASH, g–SMASH), noise statistics are Rician, as it has been widely reported and exploited previously. On the other hand, the analysis in Chapter 3 was centered on non–central Chi statistics, which are much less familiar to the diffusion imaging community, and have not been explored thoroughly. This distribution may be used to describe noise in multiple coil arrays without **k**–space subsampling if SoS is used. Although it is not exact when subsampling is performed, it has been shown that, in fact, the non–central Chi model is an accurate representation of noise, at least for the GRAPPA algorithm for pMRI. The popularity this algorithm is gaining justifies the importance of this distribution. The study in Chapter 3 was mainly focused on GRAPPA. However, all existing pMRI algorithms share a number of features that make them similar. Moreover, the reconstructions performed in one of the signal spaces studied (the **k**–space or the **x**–space) can be often mapped to other space, so any possible pMRI technique to be introduced will also show very similar characteristics to GRAPPA or SENSE. This allows to conclude that the Rician/non–central Chi model is general enough to represent most of the MRI scenarios.

Chapter 4 was centered on the analysis of the propagation of noise from the diffusion data sets to the estimated architectures in the white matter. This study was necessarily limited to a single case study, so that DTI based on non–linear estimation, and non–Gaussian diffusion schemes (i.e. HARDI estimation) were obviated. Despite this limitation, the technique studied (DTI with linear estimation) is the most common approach in current studies based on diffusion imaging. The significance of the bias introduced by Rician noise in diffusion related parameters has been previously reported in a number of empirical studies. The conclusion of the work here presented is that, for non–central Chi signals, the impact is even greater than in the common Rician scenario. In many cases, the bias in the estimation of tensor components may be critical; yet, in these situations, increasing the number of gradient directions is not advisable, since the bias is not reduced. In fact, the results in Figs. 4.2 and 4.3 provide valuable guidelines to decide whether increasing the number of gradient directions or the number of repetitions (NEX) is more convenient. It is worth to stress that this study was analytical, so the conclusions inferred show general validity withstanding any dependency on a particular kind of data.

There exists a wide variety in the properties of diffusion data sets, and also in the techniques for the description of neural architectures. As a consequence, including the modeling of noise in the estimation of fiber populations would not allow a flexible design able to account for all possible scenarios. The solution given in this work, as proposed in Chapter 5, was to decouple both problems by means of image denoising techniques: diffusion data sets were filtered accounting for their statistical properties, and the resulting *noise–free* data could be studied regardless of the scanning protocol. The estimation of neural fibers usually includes a regularization of diffusion signals (by means of LS), which imposes the condition that the input diffusion volume has to be contaminated with additive, symmetrically distributed noise. Therefore, all filtering techniques described were designed to *remove* the bias in the images due to the Rician (or non–central Chi) nature of noise. The other important conclusion from Chapter 5 is the convenience of developing filtering strategies specific for DWI data. Diffusion volumes are not only multichannel signals, but instead the correlation between different gradient directions, inherent to the diffusion model, provides very rich *a priori* information which should be exploited. A number of novel techniques in this sense were presented, showing the benefits of this approach compared to state of the art filtering techniques.

The main contribution of this dissertation was developed in Chapters 6 and 7. The aim was to find probabilistic estimators of fiber populations in the white matter of the brain. The importance of probabilistic information in image processing in general, and for this task in particular, is indubitable: probability densities provide information not only on the orientation of fiber bundles, but also on the uncertainty of the estimation performed. Together with this important issue, two additional problems inherent to any HARDI estimator were detected: first, HARDI data sets provide incomplete information on the diffusion of water molecules, so it is often necessary to make certain assumptions on the attenuation signal to *complete* this information. Second, the need for parametric models (like mixtures of Gaussians, continuous mixtures of Gaussians, and others) in some estimation techniques may reduce the generality of these approaches. Yet, in many cases, these models bind to estimate some modeling information, such as deconvolution kernels, which in turn has to be considered constant for all the extent of the white matter. These three issues are not independent, but they are closely related: in Q–Ball imaging, neither any assumption on the diffusion signal nor a prior model is required, but this estimator inherently compels to relinquish the estimation of probabilistic information. With spherical deconvolution, any kind of orientation information may be inferred without the assumption of any particular behavior of the attenuation signal, at the expense of considering a prior model for fiber populations.

The previous discussion is a remarkable conclusion derived from the analysis carried out in Chapters 6 and 7. This same discourse allowed to derive various novel estimators of neural architectures: the OPDT introduced in Chapter 6 avoided the main limitation of Q–Balls by including the Jacobian term in the radial integral. It was done at the expense of including an assumption on the diffusion signal. Withal, this assumption was weak enough to prevent the appearance of important modeling errors. Although the OPDT was shown to be more accurate than a number of state of the art approaches, gathering all the advantages of non–parametric models, it still showed a number of limitations due to the use of the FRT. The most critical was the theoretical limit it imposed to the use of high b–values. In Chapter 7, these limitations of the FRT, together with some possible solutions, were explored by varying the assumptions and modeling of the attenuation signal. The most meaningful result in this sense was the introduction of the cOPDT, which allowed to overcome the main limitations of the OPDT without any additional assumption. What is more important, this new estimator outperforms the OPDT in many situations. Together with the OPDT and the cOPDT, some additional estimators have been proposed which may me useful in different situations depending on the anatomies being imaged, the quality of the diffusion data sets (noise power), and the imaging parameters (b–value). A general conclusion inferred from Chapter 7 is the convenience of probabilistic estimation: besides its inherent benefits (characterization of uncertainty), it was shown that in many cases this kind of estimators drive to more accurate estimates. The main limitation in this study was in the numerical validation of the accuracy of the estimation of fiber populations. The great complexity of neural architectures, together with the complicated physical processes driving diffusion, makes it unfeasible to construct realistic phantoms mimicking true structures in the white matter. As a consequence, the validation had to be made over simplified scenarios with mixtures of Gaussians.

As a complementary study to that carried out in Chapter 4, the propagation of noise for the estimators presented in Chapters 6 and 7 was empirically studied. It was shown that, in most of cases, the estimators are robust enough to noise, at least for the most common Rician case. However, as shown in Chapter 8, all the estimators proposed (and any other previously reported) highly benefit from the filtering techniques discussed in Chapter 5. Chapter 8 showed extensive qualitative results on true diffusion data sets with different characteristics, proving the usefulness of the techniques described throughout this dissertation. In particular, the ability to distinguish complex neural architectures beyond the information provided by DTI was evidenced. The estimation techniques described, together with the denoisng based on unbiased filtering, were able to notably improve the results achieved with DTI even for data sets with moderate b–values, which in principle are not specifically intended for this kind of processing.

9.1 Contributions

Although all the contributions presented in this dissertation have been highlighted at the beginning of the corresponding Chapter, it is useful to include here a brief summary of them:

- In Chapter 3, the derivation of the spatial model for the variance of noise in each receiving coil following GRAPPA reconstruction. The analysis of the conditions under which a non–central Chi distribution may be assumed for this algorithm, and the study of its stationary behavior. The formalization of the general quasi–stationary non–central Chi framework for the most common single coil, multiple coil, and pMRI protocols.

- In Chapter 4, the derivation of closed–form expressions for the mean and variance of the log–Rician and log–non–central Chi models, together with their operative linear approximations. The analytical study of the propagation of noise from DWI components to components of the diffusion tensor. The study of the relation between the bias and the variance in the estimation of these components, and the analysis of the impact of the number of gradients and the number of expositions in these parameters, as a function of the SNR and the number of receiving coils.

- In Chapter 5, the filtering methodology based on the exploitation of joint information. The filters developed following such methodology: joint LMMSE, LMMSE–N and UNLM–N. The design of phantom data for the quantitative evaluation of filtering performance.

- In Chapter 6, the study on the relations between different orientation information, assumptions on the attenuation signal, and prior models for diffusion. The computation of true probability densities based on the estimation of the Laplacian operator and the computation of the FRT. The development and validation of the OPDT.

- In Chapter 7, the new insights into the relations between the scanned signals in the $\mathbf{q}$–space and the diffusion propagators (and the corresponding orientation information) in the $\mathbf{R}$–space. The use of Stoke's theorem as an alternative to the raw FRT, avoiding at the same time the need for more restrictive assumptions. The development and testing of a number of new estimators: cQ–Balls, pQ–Balls, the cOPDT, and the pOPDT[1].

9.2 Future lines

The techniques and algorithms presented in this thesis do not comprise closed–form solutions to all the current trends in diffusion imaging. On the contrary, these developments offer a collection of new techniques to tackle a number of challenging problems. Besides, the new methods themselves drive to the statement of new problems of interest. As a consequence, the work here presented may be the origin of several lines for future research. Among these directions, the most evident are the following ones:

- In the GRAPPA reconstruction algorithm (and other pMRI protocols related to this one), it was shown that the variance of noise for each coil is different, so a non–central Chi model cannot be strictly assumed. This condition may be relaxed if the variances of noise in all coils are similar. In the same way, the non–central Chi model requires that the noisy processes at each coil are independent. However, two sources of correlation may be found in real data sets: first, the signals acquired by each coil are not completely independent, and correlations may appear. This correlation is usually negligible for human and animal subjects, so it is rarely an issue. On the other hand, the interpolation of the $\mathbf{k}$–space for each coil takes into account all the remaining coils, which clearly introduces a certain correlation. Although it is expected to be low, it would be interesting to quantify this correlation and its effect on the non–central Chi modeling.

---

[1] Two of these estimators, the pOPDT and pQ–Balls, have also been proposed in two recent papers, [Aga09] and [CR09], respectively. However, the development presented in this dissertation was carried out in parallel, and yet the derivations are different.

- The study of the propagation of noise in Chapter 4 is based on the study of the properties of the transition matrix driving from the **q**–space to the **R**–space. For some HARDI estimators, like Q–Balls (and its variants), the cOPDT or the pOPDT, the estimation is performed in terms of a corresponding transition matrix, possibly operating on a contrast–corrected attenuation signal. Therefore, two tasks can be identified which could allow a similar analytical study: first, the characterization of noise in the contrast–corrected signal (a linear approximation will be necessary), and second, the study of the properties of the corresponding transition matrices.

- The validation of the filtering techniques for Rician noise has been quite extensive due to its importance in current protocols. All the derivations for the non–central Chi case have been provided as well, but they have not been properly validated. This is related to the lack of real data of this kind (see Section 1.4). The extensive validation of the proposed methods for multiple coil and pMRI scenarios (including the possible non stationary behavior of the noise power) is another interesting topic.

- The most attractive directions offered by the framework in this thesis are those related to the potential applications of the methods derived. In particular, the glyph representations in Chapter 8 provide valuable information on the architecture and connectivity of the white matter. Notwithstanding, they present an intrinsic limitation of diffusion data sets, which is the restriction to radially symmetric density functions which often do not match a realistic scenario. The posterior regularization of the field of probability densities (i.e. the regularization of the **R**–space) allows to get rid of this restriction, yielding richer representations of water diffusion known as super–resolution diffusion imaging. This line has begun to be explored recently with promising results [Bar08, NG08]. Besides, it is evident that the resolution of fiber crossings, bending, or other complex architectures may highly benefit fiber tracking algorithms, which may get lost in this kind of voxels with conventional DTI. These new approaches to tractography, which are gaining an increasing interest as well [Des09], could be even combined with the previous regularization/super–resolution approaches to achieve even more accurate results. Finally, the nice property of the OPDT, the cOPDT, and the pODPT of being estimators of true probability densities could be of high interest to trace connectivity maps of the white matter, including uncertainty management.

- The advantages of the filtering techniques described over conventional methods have been evidenced in many cases (above all, in HARDI scenarios). It is therefore interesting to test the impact of this improved processing in well established methodologies for DTI and HARDI. Moreover, the use of joint information can be extended to other filtering approaches previously reported.

- As a final remark, the adequate software integration of the methods presented in this thesis, above all those related to the estimation of fiber populations, is of capital importance for their dissemination and improvement.

A

Statistical parameters of log–Rician and log–non–central Chi distributions

In this appendix, the mean and variance of logarithmic signals obtained from the CMS are computed. Although Rician statistics are a particular case of non–central Chi when only one degree of freedom ($L = 1$) is considered, it is useful to derive the corresponding expressions separately, since the particularization of the results for the non–central Chi model, when $L = 1$ is taken, is not trivial.

A.1 Log–Rician distribution

Log–Rician statistics are obtained when the logarithm of a Rician–distributed random variable is computed:

$$D = \log(M) = \log\left(\sqrt{(A_c + n_c)^2 + (A_s + n_s)^2}\right), \tag{A.1}$$

where n_c and n_s are the independent, Gaussian distributed, real and imaginary parts of a complex Gaussian random variable, each of them with zero mean and standard deviation σ_n^2. The real and imaginary parts of the signal of interest, A_c and A_s, are also assumed to be independent from noise. The corresponding PDF may be written in the form:

$$f_D(t) = \frac{e^{2t}}{\sigma_n^2} \exp\left(-\frac{e^{2t} + A^2}{2\sigma_n^2}\right) I_0\left(\frac{Ae^t}{\sigma_n^2}\right), \tag{A.2}$$

where $t \in (-\infty, +\infty)$, and $A = \sqrt{A_c^2 + A_s^2}$ is the modulus of the signal of interest. I_0 is the modified Bessel function of the first kind and index 0. The mean of this distribution is computed directly using its definition:

$$E\{D\} = \int_{-\infty}^{\infty} t \frac{e^{2t}}{\sigma_n^2} \exp\left(-\frac{e^{2t} + A^2}{2\sigma_n^2}\right) I_0\left(\frac{Ae^t}{\sigma_n^2}\right) dt. \tag{A.3}$$

This integral does not admit a simple, closed–form primitive. Instead, the power series expansions of I_L is used [Abr72]:

$$I_L(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k+L}}{k!\Gamma(k+L+1)}, \tag{A.4}$$

where $\Gamma(z)$ stands for Euler's gamma function. Casting the previous expression (for $L=0$) in eq. (A.3), it follows:

$$\begin{aligned} E\{D\} &= \int_{-\infty}^{\infty} t\frac{e^{2t}}{\sigma_n^2}\exp\left(-\frac{e^{2t}+A^2}{2\sigma_n^2}\right)\sum_{k=0}^{\infty}\frac{1}{(k!)^2}\left(\frac{Ae^t}{2\sigma_n^2}\right)^{2k} dt \\ &= \sum_{k=0}^{\infty}\frac{A^{2k}\exp\left(-\frac{A^2}{2\sigma_n^2}\right)}{(k!)^2(2\sigma_n^2)^{2k}\sigma_n^2}\int_{-\infty}^{\infty} t\ e^{(2+2k)t}\ \exp\left(-\frac{e^{2t}}{2\sigma^2}\right)dt. \end{aligned} \tag{A.5}$$

The integral in the previous equation may be expressed in terms of the polygamma function of order 0, $\psi(z)$, defined as the first derivative of the logarithm of Euler's Gamma function [Abr72]. Hence, eq. (A.5) becomes:

$$\begin{aligned} E\{D\} &= \sum_{k=0}^{\infty}\frac{A^{2k}\exp\left(-\frac{A^2}{2\sigma_n^2}\right)}{(k!)^2(2\sigma_n^2)^{2k}\sigma_n^2}\left((2\sigma_n^2)^{k+1}\frac{k!}{4}\left(\log(2\sigma_n^2)+\psi(k+1)\right)\right) \\ &= \frac{1}{2}\exp\left(-\frac{A^2}{2\sigma_n^2}\right)\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{A^2}{2\sigma_n^2}\right)^k\left(\log(2\sigma_n^2)+\psi(k+1)\right) \\ &= \frac{1}{2}\log(2\sigma_n^2)+\frac{1}{2}\exp\left(-\frac{A^2}{2\sigma_n^2}\right)\sum_{k=0}^{\infty}\frac{\psi(k+1)}{k!}\left(\frac{A^2}{2\sigma_n^2}\right)^k. \end{aligned} \tag{A.6}$$

The summation in the previous equation can be evaluated with the help of a software for symbolic calculus. The following equalities can be obtained this way:

$$\begin{aligned} E\{D\} &= \frac{1}{2}\log(2\sigma_n^2)+\frac{1}{2}\Gamma^{(l)}\left(0,\frac{A^2}{2\sigma_n^2}\right)+\frac{1}{2}\log\left(\frac{A^2}{2\sigma_n^2}\right) \\ &= \frac{1}{2}\Gamma^{(l)}\left(0,\frac{A^2}{2\sigma_n^2}\right)+\log(A), \end{aligned} \tag{A.7}$$

with $\Gamma^{(l)}(a,b)$ the (lower) incomplete Gamma function. The resulting expression is relatively simple, and has been previously reported in [Sal05].

It remains to compute the variance of the log–Rician random variable. To that end, the second order moment is calculated first, and the variance will be obtained by subtracting the squared mean from this value. For the second order moment, the definition is used over again:

$$\begin{aligned} E\{D^2\} &= \int_{-\infty}^{\infty} t^2\frac{e^{2t}}{\sigma_n^2}\exp\left(-\frac{e^{2t}+A^2}{2\sigma_n^2}\right)I_0\left(\frac{Ae^t}{\sigma_n^2}\right)dt \\ &= \int_{-\infty}^{\infty} t^2\frac{e^{2t}}{\sigma_n^2}\exp\left(-\frac{e^{2t}+A^2}{2\sigma_n^2}\right)\sum_{k=0}^{\infty}\frac{1}{(k!)^2}\left(\frac{Ae^t}{2\sigma_n^2}\right)^{2k} dt \\ &= \sum_{k=0}^{\infty}\frac{A^{2k}\exp\left(-\frac{A^2}{2\sigma_n^2}\right)}{(k!)^2(2\sigma_n^2)^{2k}\sigma_n^2}\int_{-\infty}^{\infty} t^2\ e^{(2+2k)t}\ \exp\left(-\frac{e^{2t}}{2\sigma_n^2}\right)dt \\ &= \sum_{k=0}^{\infty}\frac{A^{2k}\exp\left(-\frac{A^2}{2\sigma_n^2}\right)}{(k!)^2(2\sigma_n^2)^{2k}\sigma_n^2}\left(\frac{(2\sigma_n^2)^{1+k}}{8}\Gamma(1+k)\left((\log(2\sigma_n^2)+\psi(1+k))^2+\psi^{1)}(1+k)\right)\right) \\ &= \frac{1}{4}\log^2\left(2\sigma_n^2\right)+\frac{1}{2}\log\left(2\sigma_n^2\right)\left(\Gamma^{(l)}\left(0,\frac{A^2}{2\sigma_n^2}\right)+\log\left(\frac{A^2}{2\sigma_n^2}\right)\right)+\frac{1}{4}\widetilde{N}_1\left(\frac{A^2}{2\sigma_n^2}\right), \end{aligned} \tag{A.8}$$

where now $\psi^{1)}(z)$ is the first derivative of $\psi(z)$ (or the second derivative of $\log\Gamma(z)$), and $\widetilde{N}_L(z)$ stands for the summation:

$$\widetilde{N}_L(z) = e^{-z}\sum_{k=0}^{\infty}\frac{z^k}{k!}\left(\psi(k+L)^2+\psi^{1)}(k+L)\right), \tag{A.9}$$

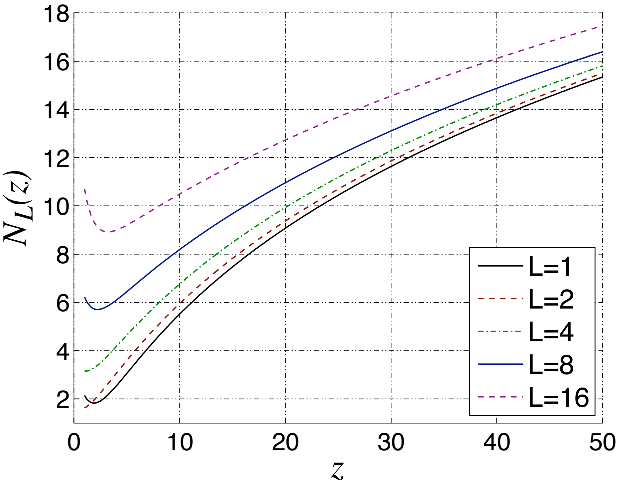


Figure A.1: Behavior of the function $\widetilde{N}_L(z)$ for different values of L and positive z.

which cannot be otherwise expressed in terms of elemental functions. This function is sketched in Fig. A.1 for different values of L. Finally, the variance of the log–Rician variable can be computed subtracting eq. (A.7) squared from eq. (A.8):

$$\mathrm{Var}\{D\} = E\{D^2\} - E^2\{D\} = \frac{1}{4}\left(\widetilde{N}_1\left(\frac{A^2}{2\sigma_n^2}\right) + \left(\Gamma^{(l)}\left(0, \frac{A^2}{2\sigma_n^2}\right) + \log\left(\frac{A^2}{2\sigma_n^2}\right)\right)^2\right). \quad (A.10)$$

A.2 Log–non–central Chi distribution

Log–non–central Chi statistics are obtained when the logarithm of a non–central Chi distributed variable is computed:

$$D_L = \log(M_L) = \log\left(\sqrt{\sum_{l=1}^{L}(A_{l_c} + n_{l_c})^2 + (A_{l_s} + n_{l_s})^2)}\right), \quad (A.11)$$

where an additional subindex, l, has to be used to denote each of the L complex signals of interest, $A_{l_c} + jA_{l_s}$, which are combined. In this case, all the complex, Gaussian distributed noisy processes $n_{l_c} + jn_{l_s}$ are assumed to be independent, and independent from the signal of interest. In this case, the PDF has the form:

$$f_{D_L}(t) = \frac{A_L^{1-L}}{\sigma_n^2} e^{(L+1)t} \exp\left(-\frac{e^{2t} + A_L^2}{2\sigma_n^2}\right) I_{L-1}\left(\frac{A_L e^t}{\sigma_n^2}\right), \quad (A.12)$$

with $t \in (-\infty, +\infty)$. I_{L-1} is the modified Bessel function of the first kind and order $L-1$, and A_L is defined as the combination of all the signals of interest when no noise is present:

$$A_L^2 = \sum_{l=1}^{L}\left(A_{l_c}^2 + A_{l_s}^2\right). \quad (A.13)$$

The mean of the new random variable can be computed using its definition, and substituting the Bessel function by its power series expansion:

$$\begin{aligned}
E\{D_L\} &= \int_{-\infty}^{\infty} t\frac{A_L^{1-L}}{\sigma_n^2} e^{(L+1)t}\exp\left(-\frac{e^{2t}+A_L^2}{2\sigma_n^2}\right) I_{L-1}\left(\frac{A_L e^t}{\sigma_n^2}\right) dt \\
&= \int_{-\infty}^{\infty} t\frac{A_L^{1-L}}{\sigma_n^2} e^{(L+1)t}\exp\left(-\frac{e^{2t}+A_L^2}{2\sigma_n^2}\right) \sum_{k=0}^{\infty}\frac{1}{k!\Gamma(L+k)}\left(\frac{A_L e^t}{2\sigma_n^2}\right)^{2k+L-1} dt \\
&= 2\sum_{k=0}^{\infty}\frac{A_L^{2k}\exp\left(-\frac{A_L^2}{2\sigma_n^2}\right)}{k!\Gamma(L+k)(2\sigma_n^2)^{L+2k}}\int_{-\infty}^{\infty} te^{2(L+1)t}\exp\left(-\frac{e^{2t}}{2\sigma_n^2}\right) dt. \qquad \text{(A.14)}
\end{aligned}$$

The previous integral can also be expressed in terms of the polygamma function:

$$\begin{aligned}
E\{D_L\} &= \frac{1}{2}\exp\left(-\frac{A_L^2}{2\sigma_n^2}\right)\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{A_L^2}{2\sigma_n^2}\right)^k \left(\log(2\sigma_n^2)+\psi(k+L)\right) \\
&= \frac{1}{2}\log\left(2\sigma_n^2\right)+\frac{1}{2}\frac{A_L^2}{2L\sigma_n^2}{}_2F_2\left(1,1:2,1+L;-\frac{A_L^2}{2\sigma_n^2}\right)+\frac{1}{2}\psi(L), \qquad \text{(A.15)}
\end{aligned}$$

where ${}_2F_2$ is the confluent hypergeomteric function of the first kind. The latter expression can be obtained using software tools for symbolic calculus.

For the second order moment:

$$\begin{aligned}
E\{D_L^2\} &= \int_{-\infty}^{\infty} t^2\frac{A_L^{1-L}}{\sigma_n^2} e^{(L+1)t}\exp\left(-\frac{e^{2t}+A_L^2}{2\sigma_n^2}\right) I_{L-1}\left(\frac{A_L e^t}{\sigma_n^2}\right) dt \\
&= \int_{-\infty}^{\infty} t^2\frac{A_L^{1-L}}{\sigma_n^2} e^{(L+1)t}\exp\left(-\frac{e^{2t}+A_L^2}{2\sigma_n^2}\right) \sum_{k=0}^{\infty}\frac{1}{k!\Gamma(L+k)}\left(\frac{A_L e^t}{2\sigma_n^2}\right)^{2k+L-1} dt \\
&= 2\sum_{k=0}^{\infty}\frac{A_L^{2k}\exp\left(-\frac{A_L^2}{2\sigma_n^2}\right)}{k!\Gamma(L+k)(2\sigma_n^2)^{L+2k}}\int_{-\infty}^{\infty} t^2e^{2(L+1)t}\exp\left(-\frac{e^{2t}}{2\sigma_n^2}\right) dt \\
&= \frac{1}{4}\exp\left(-\frac{A_L^2}{2\sigma_n^2}\right)\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{A_L^2}{2\sigma_n^2}\right)^k \left((\psi(k+L)^2+\log(2\sigma_n^2))^2+\psi^{1)}(k+L)\right) \\
&= \frac{1}{4}\left(\log(2\sigma_n^2)^2+2\log\left(2\sigma_n^2\right)\left(\psi(L)+\frac{A_L^2}{2L\sigma_n^2}{}_2F_2\left(1,1:2,1+L;-\frac{A_L^2}{2\sigma_n^2}\right)\right)\right) \\
&\quad +\frac{1}{4}\widetilde{N}_L\left(\frac{A_L^2}{2\sigma_n^2}\right). \qquad \text{(A.16)}
\end{aligned}$$

Finally, the variance can be computed as:

$$\begin{aligned}
\mathrm{Var}\{D_L\} &= \frac{1}{4}\left(\widetilde{N}_L\left(\frac{A_L^2}{2\sigma_n^2}\right)-2\log\left(2\sigma_n^2\right)\frac{A_L^2}{2L\sigma_n^2}{}_2F_2\left(1,1:2,1+L;-\frac{A_L^2}{2\sigma_n^2}\right)\right. \\
&\quad \left.-\left(\psi(L)+\frac{A_L^2}{2L\sigma_n^2}{}_2F_2\left(1,1:2,1+L;-\frac{A_L^2}{2\sigma_n^2}\right)\right)^2\right). \qquad \text{(A.17)}
\end{aligned}$$

$\mathcal{B}$

On the computation and inversion of covariance matrices

In the LMMSE approach, each noise–free channel A_i^2 is modeled as the realization of a random variable, whose mean and variance have to be estimated (only first and second order moments are required since the estimator is linear) [Kay93]. The idea behind Bayesian estimation (and therefore behind LMMSE) is to include the prior knowledge on the behavior of A_i^2 to improve the estimation. In the case of DWI data sets, the general relation given by eq. (2.39) represents a highly valuable *a priori* information: once the neural fiber structure described by $D(b,\mathbf{g})$ is known, the value of any A_i^2 may be inferred without any uncertainty from the value of any baseline A_β^2. With this consideration, only the uncertainty in the knowledge of A_β^2 has to be modeled, since all other channels are fully correlated with the baseline through $D(b,\mathbf{g}_i)$. The computation of $\mathbf{C}_{A^2M^2}$ may be done as follows:

$$\begin{aligned}
\{\mathbf{C}_{A^2M^2}\}_{ij} &= E\left\{\left(A_i^2-\langle A_i^2\rangle\right)\left(M_j^2-\langle M_j^2\rangle\right)\right\}\\
&= E\left\{\left(A_i^2-\langle A_i^2\rangle\right)\left(A_j^2+2A_jn_{c,j}+n_{c,j}^2+n_{s,j}^2-\langle A_j^2\rangle-2\sigma_n^2\right)\right\}\\
&= E\left\{\left(A_i^2-\langle A_i^2\rangle\right)\left(A_j^2-\langle A_j^2\rangle\right)\right\}=E\left\{A_i^2A_j^2\right\}-\langle A_i^2\rangle\langle A_j^2\rangle\\
&= \left(E\left\{A_\beta^4\right\}-\langle A_\beta^2\rangle^2\right)\exp\left(-bD(b,\mathbf{g}_i)\right)\exp\left(-bD(b,\mathbf{g}_j)\right)\\
&= \frac{\langle A_\beta^4\rangle-\langle A_\beta^2\rangle^2}{\langle A_\beta^2\rangle^2}\langle A_\beta^2\rangle^2\exp\left(-bD(b,\mathbf{g}_i)\right)\exp\left(-bD(b,\mathbf{g}_j)\right)\\
&= \frac{\langle A_\beta^4\rangle-\langle A_\beta^2\rangle^2}{\langle A_\beta^2\rangle^2}\langle A_i^2\rangle\langle A_j^2\rangle=\varsigma\langle A_i^2\rangle\langle A_j^2\rangle, \qquad \text{(B.1)}
\end{aligned}$$

which does not depend on $D(b,\mathbf{g}_i)$, so the filtering does not depend on the particular diffusion anatomy. The computation of the components of $\mathbf{C}_{M^2M^2}$ is more tedious, but straightforward as well. The expressions in eq. (5.15) are immediately obtained identifying corresponding terms. Note that ς is computed on the baseline β, but might be computed as well from any other gradient image of the DWI data set. However, baselines show a higher SNR and therefore are more adequate. In practice, the averaged value of ς for all baselines is used in order to reduce the variability due to the computation of the 4–th order moment. To invert $\mathbf{C}_{M^2M^2}$, let the SNR be high enough so that: $\sigma_n^2 \ll \langle A_i^2\rangle$. The following approximation might be thought of:

$$\mathbf{C}_{M^2M^2}\simeq\widetilde{\mathbf{C}}_{M^2M^2}=\varsigma\langle\mathbf{A}^2\rangle\langle\mathbf{A}^2\rangle^T+4\sigma_n^2\mathrm{diag}\left(\langle\mathbf{A}^2\rangle\right), \qquad \text{(B.2)}$$

with the advantage that the inverse of $\widetilde{\mathbf{C}}_{M^2M^2}$ may be computed in a closed form using Woodbury's identity [Kay93]:

$$\widetilde{\mathbf{C}}^{-1}_{M^2M^2} = \eta \mathbf{1}\mathbf{1}^T + \mathrm{diag}(\mathbf{e}), \tag{B.3}$$

where η is a scalar and $\mathbf{1}$ and $\mathbf{e}$ are $Z \times 1$ vectors defined as:

$$\begin{aligned} \eta &= -\left(4\sigma_n^2 \left(\frac{4\sigma_n^2}{\varsigma} + \sum_{i=0}^{L} \langle A_i^2 \rangle \right)\right)^{-1}; \\ \mathbf{1}_i &= 1; \\ \mathbf{e}_i &= \left(4\sigma_n^2 \langle A_i^2 \rangle\right)^{-1}, \end{aligned} \tag{B.4}$$

for $i = 0, \ldots, Z-1$. It is trivial to show that $\widetilde{\mathbf{C}}_{M^2M^2}$ is positive definite. Indeed, the following property holds:

$$\text{If } \langle A_i^2 \rangle > \sigma_n^2, \forall i \quad \Rightarrow \quad \lambda^j > 4\sigma_n^4, \tag{B.5}$$

for any eigenvalue λ^j of $\widetilde{\mathbf{C}}_{M^2M^2}$. To prove this relation, suppose that $\mathbf{v}^j = [v_0^j, \ldots, v_{L-1}^j]^T$ is an eigenvector associated to the eigenvalue λ^j. Then:

$$\begin{aligned} \widetilde{\mathbf{C}}_{M^2M^2}\mathbf{v}^j &= \varsigma \langle \mathbf{A}^2 \rangle \langle \mathbf{A}^2 \rangle^T \mathbf{v}^j + 4\sigma_n^2 \mathrm{diag}\left(\langle \mathbf{A}^2 \rangle\right) \mathbf{v}^j = \lambda^j \mathbf{v}^j \\ &\Rightarrow \varsigma \left(\langle \mathbf{A}^2 \rangle^T \mathbf{v}^j\right) \langle A_i^2 \rangle + 4\sigma_n^2 \langle A_i^2 \rangle v_i^j = \lambda^j v_i^j, \quad \forall i. \end{aligned} \tag{B.6}$$

For each non–null component of $\mathbf{v}^j$, λ^j may be put in the form:

$$\lambda^j = \varsigma \frac{\langle \mathbf{A}^2 \rangle^T \mathbf{v}^j}{v_i^j} \langle A_i^2 \rangle + 4\sigma_n^2 \langle A_i^2 \rangle. \tag{B.7}$$

And therefore:

- If $\langle \mathbf{A}^2 \rangle^T \mathbf{v}^j \geq 0$, it has to be at least one positive component of $\mathbf{v}^j$, $v_p^j > 0$, since $\langle A_i^2 \rangle > 0$. Particularizing eq. (B.7) for v_p^j, it is evident that $\lambda^j \geq 4\sigma_n^2 \langle A_i^2 \rangle$. From the premise in eq. (B.5), $\lambda^j > 4\sigma_n^4$.
- If $\langle \mathbf{A}^2 \rangle^T \mathbf{v}^j \leq 0$, it has to be at least one negative component of $\mathbf{v}^j$, $v_n^j < 0$. Particularizing eq. (B.7) for v_n^j, it is evident that $\lambda^j \geq 4\sigma_n^2 \langle A_i^2 \rangle \Rightarrow \lambda^j > 4\sigma_n^4$.

The property in eq. (B.5) guarantees the convergence of the power series expansion of $\mathbf{C}^{-1}_{M^2M^2}$ as a function of $\widetilde{\mathbf{C}}^{-1}_{M^2M^2}$:

$$\begin{aligned} \mathbf{C}^{-1}_{M^2M^2} &= \left(\widetilde{\mathbf{C}}_{M^2M^2} + 4\sigma_n^4 \mathrm{I}_L\right)^{-1} = \left(\widetilde{\mathbf{C}}_{M^2M^2}\left(\mathrm{I}_L + 4\sigma_n^4 \widetilde{\mathbf{C}}^{-1}_{M^2M^2}\right)\right)^{-1} \\ &= \left(\mathrm{I}_L + 4\sigma_n^4 \widetilde{\mathbf{C}}^{-1}_{M^2M^2}\right)^{-1} \widetilde{\mathbf{C}}^{-1}_{M^2M^2} = \widetilde{\mathbf{C}}^{-1}_{M^2M^2} \sum_{k=0}^{\infty} \left(-4\sigma_n^4 \widetilde{\mathbf{C}}^{-1}_{M^2M^2}\right)^k, \end{aligned} \tag{B.8}$$

since $\widetilde{\mathbf{C}}^{-1}_{M^2M^2}$ commutes with any function that can be expressed as a power series of $\widetilde{\mathbf{C}}^{-1}_{M^2M^2}$ (such function will have the same eigenspace as $\widetilde{\mathbf{C}}^{-1}_{M^2M^2}$). It is straightforward to show that the term $\mathscr{O}$ in the recursive formula of eq. (5.17) represents the product of the truncated series expansion in eq. (B.8), up to order $\mathscr{O}$, by $\mathbf{M}^2 - \langle \mathbf{M}^2 \rangle$.

C

Noise estimation in Rician and non–central Chi signals

The estimation of the noise power σ_n^2 in an image is of paramount importance not only for filtering tasks, but also in new techniques for tensor estimation [Fil07, Lan07] or inference of fiber directions [Cla08]. Like most of the Rician–based filtering algorithms, noise estimation is based on the computation of sample moments of the CMS to estimate the actual moments of the underlying distributions. From them, the parameters of these distributions, and in particular σ, are inferred. A detailed description of the techniques used to estimate σ is far beyond the scope of this dissertation, so this Appendix is intended only as a brief overview of the existing methods. The most of the related literature is focused on Rician distributed signals; however, in the same way it was done in Section 5.5.1, all of them may be easily extrapolated to the non–central Chi case. Noise estimation techniques can be divided in the following categories:

- Estimators that need prior segmentation of the background. In the background $A(\mathbf{x}) = 0$, so the statistics reduce to a Rayleigh distribution, whose moments are easily computed averaging all the voxels in the detected background. This kind of techniques may fail if the background is not accurately segmented,if a large enough background is not present (for brain images, this problem is rarely an issue), or in the case strong inhomogeneities in the background exist.
- Estimators based on the background that do not need its prior segmentation. They are based as well on a Rayleigh distribution, but instead of averaging all the available voxels to estimate the moments, they are based on the computation of local moments, whose distribution is analyzed. Three approaches may be distinguished:
    - Using the mode of a given local parameter related to σ_n, computed as the maximum of the histogram.
    - Fitting a Rayleigh–related distribution to the histogram of the local feature.
    - Using Maximum Likelihood estimators of σ_n.

  In all these cases, it is assumed that the background is large enough so that the mode of the histograms corresponds to the Rayleigh voxels. However, techniques to reject the non–background voxels may benefit the estimation;

$\overline{M(\mathbf{x})}$	(Global) Sample mean of image $M(\mathbf{x})$ $\overline{M(\mathbf{x})} = \frac{1}{\lvert\Omega\rvert} \sum_{\mathbf{x}\in\Omega} M(\mathbf{x})$
$\overline{M(\mathbf{x})}_x$	Local sample local mean of image $M(\mathbf{x})$ $\overline{M(\mathbf{x})}_x = \frac{1}{\lvert\mathcal{N}_x\rvert} \sum_{\mathbf{c}\in\mathcal{N}_x} M(\mathbf{c})$ ($\mathcal{N}_x$ is a neighborhood of size $\lvert\mathcal{N}_x\rvert = N$ centered in $\mathbf{x}$)
$\mathrm{Var}\{M(\mathbf{x})\}_x$	Sample local variance of $M(\mathbf{x})$ $\mathrm{Var}\{M(\mathbf{x})\}_x = \overline{M^2(\mathbf{x})}_x - \overline{M(\mathbf{x})}_x^2$
$M(\mathbf{x_B})$	Background area of image $M(\mathbf{x})$ $\mathbf{x_B} = \{\mathbf{x} \lvert A(\mathbf{x}) = 0\}$
$M(\mathbf{x_R})$	$M(\mathbf{x})$ in the arbitrary region $\mathbf{R}$ $\mathbf{x_R} \in \mathbf{R}$
$h_I(l)$	The value of the empirical histogram of I at bin l, $l_0 \leq l \leq l_c$, computed over a total of S samples
$\mathrm{mode}\{I(\mathbf{x})\}$	Mode of the distribution of $I(\mathbf{x})$ $\mathrm{mode}\{I(\mathbf{x})\} = \arg\max_{l}\{h_I(l)\}$
L	The number of coils in multiple coils systems

Table C.1: Summary of the notation used for the estimators of the noise power reviewed in Tables C.2 and C.3.

in particular, a rough segmentation of the background is useful to minimize the effects of Rician voxels in the estimation. In this case the segmentation has to be not so accurate, so this kind of methods is the preferred whenever a large background area is available.

- Estimators based on Rician data, which have to be used if no background is present. In general they are more complicated, and yet prone to errors due to the greater uncertainty in the computation of the moments of the Rician distribution. Nevertheless, this is the only alternative when no background is present.
- Estimators in the Gaussian distributed, complex image domain. The output provided by the MRI scanner is almost always the CMS, so the complex data is rarely available. Therefore, this category falls out of the discussion.

Of course, this classification holds for non–central Chi signals, with the only difference that in the background the distribution reduces to a central Chi instead of a Rayleigh distribution. For DWI data sets, the background region is the same for all baselines and gradient images; as a consequence, the estimation may be performed in the baselines, in the gradient images, or in both of them simultaneously for the first two kinds of techniques. DWI is nowadays mainly intended for the brain (a Rayleigh/central Chi background is always present), so the third group of techniques is of less interest. In what follows, the notation in Table C.1 is used.

Table C.2 shows the expressions of some estimators for the Rician case, following the classification outlined above. For the case of DWI images of the brain, the

Estimator				References
Rician distributed signals	Rayleigh model	Bck. segm.	$\widehat{\sigma_n^2} = \frac{1}{2}\overline{M^2(\mathbf{x_B})}$	[Now99, Sij98]
			$\widehat{\sigma_n} = \sqrt{\frac{2}{\pi}}\overline{M(\mathbf{x_B})}$	[Sij98]
		No need prior segmentation of the background	$\widehat{\sigma_n} = \text{mode}\{M(\mathbf{x})\}$	[Bru93]
			$\widehat{\sigma_n^2} = \frac{1}{2}\text{mode}\{\overline{M^2(\mathbf{x})}_x\}$	[AF08a]
			$\widehat{\sigma_n} = \sqrt{\frac{N}{2N-1}}\text{mode}\left\{\sqrt{\overline{M^2(\mathbf{x})}_x}\right\}$	[AF09]
			$\widehat{\sigma_n} = \sqrt{\frac{2}{\pi}}\text{mode}\{\overline{M(\mathbf{x})}_x\}$	[AF08a]
			$\widehat{\sigma_n^2} = \frac{2}{4-\pi}\text{mode}\{\text{Var}\{M(\mathbf{x})\}_x\}$	[AF08a]
			$\widehat{\sigma_n} = \arg\min_{\sigma,K}\sum_{l=l_0}^{l_c}\left(h_M(l) - K\frac{l}{\sigma^2}e^{-\frac{l^2}{2\sigma^2}}\right)^2$	[Bru93]
			$\widehat{\sigma_n} = \arg\min_{\sigma,K}\sum_{l=l_0}^{l_c}\left(h_{\overline{M^2}}(l) - K\frac{l^{N-1}N^N}{(2\sigma^2)^N\Gamma(N)}e^{-\frac{lN}{2\sigma^2}}\right)^2$	[AF09]
			$\widehat{\sigma_n} = \arg\min_{\sigma,K}\sum_{l=l_0}^{l_c}\left(h_{\overline{M}}(l) - K\frac{l^{2N-1}N^N}{2^{N-1}b^N\Gamma(N)}e^{-\frac{l^2N}{2b}}\right)^2$	[AF09]
			$\widehat{\sigma_n} = \arg\min_{\sigma,K}\sum_{l=l_0}^{l_c}\left(h_{\sqrt{\overline{M^2}}}(l) - K\frac{l^{2N-1}N^N}{2^{N-1}\sigma^{2N}\Gamma(N)}e^{-\frac{l^2N}{2\sigma^2}}\right)^2$	[AF09]
			$\widehat{\sigma_n} = \arg\min_{\sigma}\left[S\log\left(e^{-\frac{l_0^2}{2\sigma^2}} - e^{-\frac{l_c^2}{2\sigma^2}}\right) - \sum_{i=1}^{c}h_M(l_i)\log\left(e^{-\frac{l_{i-1}^2}{2\sigma^2}} - e^{-\frac{l_i^2}{2\sigma^2}}\right)\right]$	[Sij07]
			$\widehat{\sigma_n} = \arg\min_{\sigma}\left[S\log\left(\Gamma\left(N+1, l_0\frac{N}{2\sigma^2}\right) - \Gamma\left(N+1, l_c\frac{N}{2\sigma^2}\right)\right) - \sum_{i=1}^{c}h_{\overline{M^2}}(l_i)\log\left(\Gamma\left(N+1, l_{i-1}\frac{N}{2\sigma^2}\right) - \Gamma\left(N+1, l_i\frac{N}{2\sigma^2}\right)\right)\right]$	[AF09]
			$\widehat{\sigma_n} = \arg\min_{\sigma}\left[S\log\left(\Gamma\left(N, \frac{l_0^2N}{2\sigma^2}\right) - \Gamma\left(N, \frac{l_c^2N}{2\sigma^2}\right)\right) - \sum_{i=1}^{c}h_{\sqrt{\overline{M^2}}}(l_i)\log\left(\Gamma\left(N, \frac{l_{i-1}^2N}{2\sigma^2}\right) - \Gamma\left(N, \frac{l_i^2N}{2\sigma^2}\right)\right)\right]$	[AF09]
	Rician model		$\widehat{\sigma_n^2} = \text{mode}\{\text{Var}\{M(\mathbf{x})\}_x\}$	[AF08a]

Table C.2: Survey of noise estimators for Rician distributed data. The theoretical justification for each of them, together with some details on their implementation and performance, may be found in the references provided and in [AF09].

most interesting techniques are those based on a Rayleigh distributed background (since they are more robust to the uncertainty in the computation of sample moments than those based on a general Rician distribution). However, it is desirable that the estimation does not need an accurate segmentation of the background.

Estimator				References
Non–central Chi distributed signals	Central Chi model	Bck. segm.	$\widehat{\sigma_n^2} = \frac{1}{2L}\overline{M^2(\mathbf{x_B})}$	[Con97, Die08]
			$\widehat{\sigma_n} = \frac{1}{\sqrt{2}L}\overline{M(\mathbf{x_B})}\frac{\sqrt{L}\Gamma(L)}{\Gamma\left(L+\frac{1}{2}\right)}$	[Die08]
		No need prior segmentation of the background	$\widehat{\sigma_n} = \frac{1}{\sqrt{2L-1}}\text{mode}\{M(\mathbf{x})\}$	[AF09]
			$\widehat{\sigma_n^2} = \frac{1}{2L}\text{mode}\{\overline{M^2(\mathbf{x})}_x\}$	[AF09]
			$\widehat{\sigma_n} = \frac{1}{\sqrt{2}L}\text{mode}\{\overline{M(\mathbf{x})}_x\}\frac{\sqrt{L}\Gamma(L)}{\Gamma\left(L+\frac{1}{2}\right)}$	[AF09]
			$\widehat{\sigma_n^2} = \left(2L - \frac{2\Gamma^2\left(L+\frac{1}{2}\right)}{\Gamma^2(L)}\right)^{-1}\text{mode}\{\text{Var}\{M(\mathbf{x})\}_x\}$	[AF09]
			$\widehat{\sigma_n} = \arg\min_{\sigma,K}\sum_{l=l_0}^{l_c}\left(h_M(l) - K\frac{2^{1-L}}{\Gamma(L)}\frac{l^{2L-1}}{\sigma^{2L}}e^{-\frac{l^2}{2\sigma^2}}\right)^2$	[AF09]
			$\widehat{\sigma_n} = \arg\min_{\sigma,K}\sum_{l=l_0}^{l_c}\left(h_{\overline{M^2}}(l) - K\frac{l^{NL-1}N^{NL}}{(2\sigma^2)^{NL}\Gamma(NL)}e^{-\frac{lN}{2\sigma^2}}\right)^2$	[AF09]
			$\widehat{\sigma_n} = \arg\min_{\sigma}\Big[S\log\left(\Gamma\left(L,\frac{l_0^2}{2\sigma^2}\right) - \Gamma\left(L,\frac{l_c^2}{2\sigma^2}\right)\right)$ $-\sum_{i=1}^{c}h_M(l_i)\log\left(\Gamma\left(L,\frac{l_{i-1}^2}{2\sigma^2}\right) - \Gamma\left(L,\frac{l_i^2}{2\sigma^2}\right)\right)\Big]$	[AF09]
			$\widehat{\sigma_n} = \arg\min_{\sigma}\Big[S\log\left(\Gamma\left(NL,l_0\frac{N}{2\sigma^2}\right) - \Gamma\left(NL,l_c\frac{N}{2\sigma^2}\right)\right)$ $-\sum_{i=1}^{c}h_{\overline{M^2}}(l_i)\log\left(\Gamma\left(NL,l_{i-1}\frac{N}{2\sigma^2}\right) - \Gamma\left(NL,l_i\frac{N}{2\sigma^2}\right)\right)\Big]$	[AF09]
	Non–central Chi model		$\widehat{\sigma_n^2} = \text{mode}\{\text{Var}\{M(\mathbf{x})\}_x\}$	[AF09]

Table C.3: Survey of noise estimators for non–central Chi distributed data. The theoretical justification for each of them, together with some details on their implementation and performance may be found in the references provided and in [AF09].

In this sense, the first two techniques are based on moments computed over the whole set of background voxels, so they will be highly influenced by errors in the segmentation: in particular, if Rician distributed voxels are included in the computation, σ_n will be overestimated. For the remaining estimators, any of the three groups of techniques might be used. The estimators in the first group, i.e. those based on the mode of local distributions, are especially interesting due to their simplicity. Among them, the estimator based on the mode of $\overline{M(\mathbf{x})}_x$ [AF08a] shows a nice behavior, and so it has been used in all the experiments presented in this dissertation.

Finally, Table C.3 shows the analogous results for non–central Chi signals. Practically all the estimators in Table C.2 have their corresponding counterpart in Table C.3. In particular, there is an estimator based on the mode of $\overline{M(\mathbf{x})}_x$, whose advantages have been mentioned above.

D

SH expansions and computation of some HARDI estimators

The analysis and synthesis expressions of the SH expansion given in eq. (6.16), as well as the integral defining the FRT in eq. (6.7) for this representation may be written as matrix operations [Des07]. In this Appendix this work is briefly reviewed and extended to the estimators presented in Chapters 6 and 7.

D.1 Notation

Like in [Des07], the indexation of the SH basis is changed to represent them with one single index j. For each even order l, it is easy to note that there are $H_l = 2l+1$ basis functions $\widetilde{Y}_l^m$, so the following indexation may be used:

$$\{\widetilde{Y}_j,\ j=\frac{l}{2}(l-1)\dots\frac{l}{2}(l+3)\}=\{\widetilde{Y}_l^{-l}\dots\widetilde{Y}_l^{l}\},\ l \text{ even.} \tag{D.1}$$

Note that only even order SH are considered, according to eq. (6.16). Besides, the notation l_j is used for the SH order of the corresponding basis function:

$$l_j = l:\ \frac{l}{2}(l-1)\leq j\leq\frac{l}{2}(l+3),\ l \text{ even.} \tag{D.2}$$

It is easy to note that the total number of SH basis functions for an expansion up to order $L=2\hat{L}$ is $H=\sum_{l=0}^{\hat{L}}H_{2l}=(L/2+1)(L+1)$. The set of N sampling directions of $E(\mathbf{q})$ is denoted as $\Theta=\{(\theta_1,\phi_1),(\theta_2,\phi_2),\dots,(\theta_N,\phi_N)\}\equiv\{\mathbf{g}_1,\mathbf{g}_2,\dots,\mathbf{g}_N\}$; the set of N' sampling directions for which the OPDF (or any other orientation information) is computed is $\Theta'=\{(\theta'_1,\phi'_1),(\theta'_2,\phi'_2),\ \dots,(\theta'_{N'},\phi'_{N'})\}\equiv\{\mathbf{r}_1,\mathbf{r}_2,\dots,\mathbf{r}_{N'}\}$. The $N\times 1$ vectors $\mathbf{E}$ (attenuation signals) and $\mathbf{D}$ (log–signal), the $H\times 1$ vectors $\mathbf{C}$ (SH coefficients of E or ΔE) and $\mathbf{C}'$ (the SH coefficients of the ODF or the OPDF), and the $N'\times 1$ vector $\mathbf{P}$ (the orientation function at each direction in Θ') are defined as:

$$\begin{array}{rcl}\{\mathbf{E}\}_i & = & E(q_0\mathbf{g}_i);\ \{\mathbf{D}\}_i=\log E(q_0\mathbf{g}_i);\ (N\times 1).\\ \{\mathbf{C}\}_j & = & C_j;\ \{\mathbf{C}'\}_j=C'_j;\ (H\times 1).\\ \{\mathbf{P}\}_{i'} & = & \Psi(\mathbf{r}_{i'}) \text{ or } \Phi(\mathbf{r}_{i'});\ (N'\times 1).\end{array} \tag{D.3}$$

The matrix of eigenvalues of the SH basis, $\mathbf{L}$, and the FRT matrix, $\mathscr{F}$, are $H \times H$ diagonal matrices of the form:

$$\{\mathbf{L}\}_{jj} = -l_j(l_j+1);\ \{\mathscr{F}\}_{jj} = \begin{cases} 1, & j=0 \\ 2\pi(-1)^{l_j/2}\frac{(l_j-1)!!}{l_j!!}, & j>0 \end{cases}, \tag{D.4}$$

where the standard notation for the double factorial has been used:

$$n!! = \begin{cases} 1\cdot 3\cdot 5\cdot \ldots \cdot n, & n \text{ odd} \\ 2\cdot 4\cdot 6\cdot \ldots \cdot n, & n \text{ even} \end{cases}. \tag{D.5}$$

Finally, $\mathbf{B}$ (resp. $\mathbf{B}'$) is an $N \times H$ (resp. $N' \times H$) matrix of evaluations of the H SH basis functions in the set Θ (resp. Θ'):

$$\{\mathbf{B}\}_{ij} = \widetilde{Y}_j(\theta_i,\phi_i);\ \{\mathbf{B}'\}_{i'j} = \widetilde{Y}_j(\theta'_{i'},\phi'_{i'}); \tag{D.6}$$

D.2 SH analysis and synthesis formulas

Given the noisy vector $\mathbf{E}$ of the evaluations of E over Θ, and the SH matrix $\mathbf{B}$ of the evaluations of the SH basis functions in these same directions Θ, the computation of the coefficients vector $\mathbf{C}$ may be seen as a LS problem, whose closed–form solution is well known:

$$\mathbf{BC} \simeq \mathbf{E} \Rightarrow \mathbf{C} = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{E}. \tag{D.7}$$

In order to achieve an adequate representation capability beyond common parametric models, the maximum order L of the basis functions must be high enough. However, using high SH orders translates in badly conditioned matrices $(\mathbf{B}^T\mathbf{B})$, so the LS fitting may highly amplify the noise. It is very convenient to include a regularization penalty term in the problem, typically associated to the energy of the second derivatives (Laplacian) of the function to interpolate. Due to the orthogonality of the SH basis functions, this penalty has a very simple form:

$$\mathbf{C} = \left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\mathbf{E}. \tag{D.8}$$

Alternatively, once the vector of coefficients $\mathbf{C}$ is determined, it is very easy to compute the interpolated values of the estimated functions for a different set of directions Θ', $\mathbf{E}'$:

$$\mathbf{E}' = \mathbf{B}'\mathbf{C}. \tag{D.9}$$

Note the high simplicity of these approximations compared to the actual values in eq. (6.16). Besides, the inclusion of the custom parameter λ allows the regularization of the functions to interpolate to palliate the effect of noise.

D.3 Q–Balls computation

The diffusion signal sampled in the sphere $\mathscr{S}$ can be represented in terms of its SH expansion up to order L using eq. (D.8). This means that the (regularized) function could be recovered for any spatial direction $\mathbf{r}$ evaluating eq. (6.16). Taking into account the property of SH of being eigenfunctions for this transform, the FRT of each basis function is computed multiplying it by the corresponding FRT eigenvalue. Due to its linearity, the FRT of $E(q_0\mathbf{g})$ may be expanded in the SH basis with coefficients which are the same as those of $E(q_0\mathbf{g})$ multiplied by the FRT eigenvalues, see eq. (6.17). The SH coefficients $\mathbf{C}'$ of the FRT of $E(q_0\mathbf{g})$ are:

$$\mathbf{C}' = \mathscr{F}\mathbf{C} = \mathscr{F}\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\mathbf{E}. \tag{D.10}$$

It only remains to evaluate the SH expansion at the directions in Θ' to obtain:

$$\mathbf{P} = \mathbf{B}'\mathbf{C}' = \mathbf{B}'\mathscr{F}\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\mathbf{E}. \tag{D.11}$$

Note that the whole eq. (D.11) may be precomputed for all voxels except for the product with $\mathbf{E}$, so the estimation of $\Psi(\mathbf{r})$ requires only a product of an $N' \times N$ matrix by an $N \times 1$ vector.

D.4 OPDT computation

The OPDT may be lastly represented as the FRT of the estimated Laplacian of $E(q_0\mathbf{g})$, so a similar analysis to that of the previous Section may be performed. Since the Laplacian may be divided in two different parts, one corresponding to the angular derivatives and one corresponding to the radial one, two analysis are to be performed; the linearity of the FRT will be used to join both parts. For the radial derivative, it follows from eq. (6.26) that the following approximation holds:

$$\Delta_q\mathbf{E} \simeq \frac{2}{q_0^2}\left[\mathbf{D}\cdot(3+2\mathbf{D})\cdot\mathbf{E}\right], \tag{D.12}$$

where the brackets $[\cdot]$ denote element–wise operations. A reasoning similar to that of Section D.3 yields:

$$\mathbf{C}'_q = \mathscr{F}\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\left[\frac{2}{q_0^2}\mathbf{D}\cdot(3+2\mathbf{D})\cdot\mathbf{E}\right], \tag{D.13}$$

where $\mathbf{C}'_q$ are the terms of the SH expansion of the OPDT corresponding to the radial derivative. For the angular derivatives, the property of SH of being eigenfunctions of the Laplace–Beltrami operator is exploited; the SH coefficients of $\Delta_b E(q_0\mathbf{g})$, $\mathbf{C}_b$, will be:

$$\mathbf{C}_b = \frac{1}{q_0^2}\mathbf{L}\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\mathbf{E}, \tag{D.14}$$

and the SH coefficients of its FRT, $\mathbf{C}'_b$:

$$\mathbf{C}'_b = \frac{1}{q_0^2}\mathscr{F}\mathbf{L}\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\mathbf{E}. \tag{D.15}$$

From eqs. (D.13) and (D.15) the SH coefficients of the OPDT are:

$$\begin{aligned}\mathbf{C}' &= \mathscr{F}\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\left[\frac{2}{q_0^2}\mathbf{D}\cdot(3+2\mathbf{D})\cdot\mathbf{E}\right] + \mathscr{F}\frac{1}{q_0^2}\mathbf{L}\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\mathbf{E} \\ &= \frac{1}{q_0^2}\mathscr{F}\left(\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\left[2\mathbf{D}\cdot(3+2\mathbf{D})\cdot\mathbf{E}\right] + \mathbf{L}\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\mathbf{E}\right).\end{aligned} \tag{D.16}$$

Finally, the estimator for the OPDT is written:

$$\mathbf{P} = \frac{-1}{4\pi^2 q_0^2}\mathbf{B}'\mathscr{F}\left(\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\left[2\mathbf{D}\cdot(3+2\mathbf{D})\cdot\mathbf{E}\right] + \mathbf{L}\left(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\mathbf{E}\right). \tag{D.17}$$

Note that it is not necessary to know the exact value of q_0 if $\Phi(\mathbf{r})$ is normalized, as opposed to the case of the DOT, see [Öza06, Appendix A, eq. (28)].

D.5 Other integral estimators based on the FRT

A number of novel estimators were developed in Chapter 7, whose implementations in terms of the FRT were summarized in Table 7.1. Generally speaking, these estimators may be written in terms of the FRT of a contrast enhanced function computed from the attenuation signal $E(\mathbf{q})$ (or its Laplacian). As a consequence, they are easily related to the estimators reviewed above, and so their implementations are. The corresponding expressions are given below without further comments, since their deduction is trivial from Table 7.1 taking into account the previous developments:

$$\text{cQ–Balls:} \quad \mathbf{P} = \frac{-q_0^2}{4}\mathbf{B}'\mathscr{F}\left(\mathbf{B}^T\mathbf{B}+\lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\left[\frac{1-\mathbf{E}}{\log\left(\mathbf{E}\right)}\right]. \tag{D.18}$$

$$\text{pQ–Balls:} \quad \mathbf{P} = \frac{-q_0^2}{4}\mathbf{B}'\mathscr{F}\left(\mathbf{B}^T\mathbf{B}+\lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\left[\frac{1}{\log\left(\mathbf{E}\right)}\right]. \tag{D.19}$$

$$\text{cOPDT:} \quad \mathbf{P} = \frac{-1}{16\pi^2}\mathbf{B}'\mathscr{F}\mathbf{L}\left(\mathbf{B}^T\mathbf{B}+\lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\left[\mathrm{E}_{in}\left(\mathbf{E}\right)\right]+\frac{1}{4\pi}. \tag{D.20}$$

$$\text{pOPDT:} \quad \mathbf{P} = \frac{1}{16\pi^2}\mathbf{B}'\mathscr{F}\mathbf{L}\left(\mathbf{B}^T\mathbf{B}+\lambda\mathbf{L}^2\right)^{-1}\mathbf{B}^T\left[\log\left(\mathbf{E}\right)\right]+\frac{1}{4\pi}. \tag{D.21}$$

Like in eq. (D.11), all matrix operations may be precomputed, so that at each voxel it is only necessary to compute the operation in brackets and pre–multiply the resulting vector by a fixed matrix. Besides, no additional normalization is required with the cOPDT and the pOPDT, which already have integral 1. This is an advantage over the OPDT described by eq. (D.17).

$\mathcal{E}$

Expressions of the Laplacian for the tensor model

E.1 Laplacian in Cartesian coordinates

With the notation in eqs. (2.29) and (6.20), the Stejskal–Tanner equation may be written in the form:

$$\begin{aligned} E(\mathbf{q}) \equiv E(q_1,q_2,q_3) &= \exp\left(-4\pi^2\tau\mathbf{q}^T\mathscr{D}\mathbf{q}\right) \\ &= \exp\left(-4\pi^2\tau[q_1,q_2,q_3]\mathscr{D}[q_1,q_2,q_3]^T\right), \end{aligned} \tag{E.1}$$

and so, due to the symmetry of the diffusion tensor:

$$\begin{aligned} \frac{\partial}{\partial q_1}E(\mathbf{q}) &= -4\pi^2\tau\left([1,0,0]\mathscr{D}\mathbf{q}+\mathbf{q}^T\mathscr{D}[1,0,0]^T\right)E(\mathbf{q}) \\ &= -8\pi^2\tau\ [1,0,0]\mathscr{D}\mathbf{q}\ E(\mathbf{q}). \\ \frac{\partial^2}{\partial q_1^2}E(\mathbf{q}) &= -8\pi^2\tau\left(-8\pi^2\tau\left([1,0,0]\mathscr{D}\mathbf{q}\right)^2+[1,0,0]\mathscr{D}[1,0,0]^T\right)E(\mathbf{q}). \end{aligned} \tag{E.2}$$

Computing the corresponding terms for q_2 and q_3, the Laplacian reads:

$$\Delta E = -8\pi^2\tau\left(-8\pi^2\tau\left\|\mathscr{D}\mathbf{q}\right\|^2+\operatorname{tr}(\mathscr{D})\right)E(\mathbf{q}). \tag{E.3}$$

E.2 Radial projection of the Laplacian

From eq. (2.29), taking into account that $\mathbf{q}=q\mathbf{g}$, Δ_q may be written as:

$$\begin{aligned} q^2\frac{\partial}{\partial q}E(\mathbf{q}) &= q^2\left(-8\pi^2\tau q\mathbf{g}^T\mathscr{D}\mathbf{g}\right)E(\mathbf{q}); \\ \Delta_q E(\mathbf{q}) = \frac{\partial}{\partial q}\left(q^2\frac{\partial E}{\partial q}\right) &= \left(q^2\left(-8\pi^2\tau q\mathbf{g}^T\mathscr{D}\mathbf{g}\right)^2-24\pi^2\tau q^2\mathbf{g}^T\mathscr{D}\mathbf{g}\right)E(\mathbf{q}) \\ &= 8q^2\pi^2\tau\mathbf{g}^T\mathscr{D}\mathbf{g}\left(8\pi^2\tau\mathbf{q}^T\mathscr{D}\mathbf{q}-3\right)E(\mathbf{q}). \end{aligned} \tag{E.4}$$

Note that this last equation is formally identical to the approximation in eq. (6.26). Moreover, for the tensor model, $q^2D(q,\mathbf{g})=q^2\mathbf{g}^T\mathscr{D}\mathbf{g} \Rightarrow D(q,\mathbf{g})=D(q_0\mathbf{g})=\mathbf{g}^T\mathscr{D}\mathbf{g}$ (D does not depend on q), and eq. (6.26) is the exact expression of Δ_q.

E.3 Laplace–Beltrami operator

From the relation given by eq. (6.22) between the actual expression of the Laplacian and the two corresponding projections, it immediately follows:

$$\begin{aligned}\varDelta_b E(\mathbf{q}) &= q^2\left(\varDelta - \frac{1}{q^2}\varDelta_q\right)E(\mathbf{q}) = \left(q^2\varDelta - \varDelta_q\right)E(\mathbf{q}) \\ &= -8\pi^2\tau q^2\left(-8\pi^2\tau\,\|\mathscr{D}\mathbf{q}\|^2 + \mathrm{tr}(\mathscr{D}) + \mathbf{g}^T\mathscr{D}\mathbf{g}\left(8\pi^2\tau\mathbf{q}^T\mathscr{D}\mathbf{q} - 3\right)\right)E(\mathbf{q}). \qquad \text{(E.5)}\end{aligned}$$

F

Definition and properties of the non–singular exponential integral

Consider the integral in eq. (7.22), which is reproduced here for convenience:

$$\widetilde{F_\nu^{\Phi}}(q,\xi,\nu) \simeq \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b \exp\left(-4\pi^2\tau u^2 \widetilde{D}(q_0,\xi,\nu)\right)}{u}du. \tag{F.1}$$

The ADC, $\widetilde{D}$, is a positive, bounded function defined on the unit sphere for each q_0. Therefore, the exponential term in the numerator is simply 1 for $u=0$. Without the Laplace–Beltrami operator, the resulting function has a pole of the form u^{-1}, so it is not integrable at $u=0$. To overcome this problem, the previous equation may be re–written in the form:

$$\begin{aligned}
\widetilde{F_\nu^{\Phi}}(q,\xi,\nu) &\simeq \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b\left(\exp\left(-4\pi^2\tau u^2 \widetilde{D}(q_0,\xi,\nu)\right)-1+1\right)}{u}du \\
&= \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b\left(\exp\left(-4\pi^2\tau u^2 \widetilde{D}(q_0,\xi,\nu)\right)-1\right)+\Delta_b 1}{u}du \\
&= \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b\left(\exp\left(-4\pi^2\tau u^2 \widetilde{D}(q_0,\xi,\nu)\right)-1\right)}{u}du;
\end{aligned} \tag{F.2}$$

Δ_b is a linear differential operator, so applying it to a constant yields simply 0. Besides, this operator does not depend on the radial coordinate u, since it represents precisely the angular Laplacian. Consequently, in the previous equation the order of the differential and the integral operators may be exchanged:

$$\widetilde{F_\nu^{\Phi}}(q,\xi,\nu) \simeq \frac{-1}{8\pi^2 q}\Delta_b\int_0^q \frac{\exp\left(-4\pi^2\tau u^2 \widetilde{D}(q_0,\xi,\nu)\right)-1}{u}du. \tag{F.3}$$

This step is possible because the resulting integral is now convergent. To show it, consider the power series expansion of the numerator, which is trivially related to the expansion of the exponential function:

$$\begin{aligned}
\exp\left(-4\pi^2\tau u^2 \widetilde{D}(q_0,\xi,\nu)\right)-1 &= \sum_{k=0}^{\infty}\frac{1}{k!}\left(-4\pi^2\tau u^2 \widetilde{D}(q_0,\xi,\nu)\right)^k - 1 \\
&= \sum_{k=1}^{\infty}\frac{1}{k!}\left(-4\pi^2\tau u^2 \widetilde{D}(q_0,\xi,\nu)\right)^k.
\end{aligned} \tag{F.4}$$

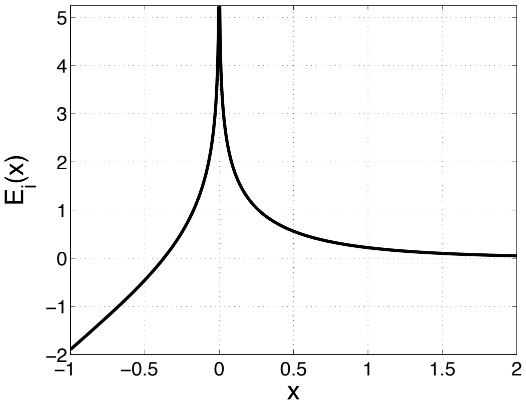


Figure F.1: The real–valued exponential integral $\mathrm{E}_i(x)$. For negative arguments, the value of the function is not unique, so the principal (real) branch has been depicted.

This power series is absolutely convergent (i.e. it converges to a bounded limit even if the modulus of each term is taken) for any particular value of $\widetilde{D}(q_0,\xi,\nu)$. From the previous expression, it is trivial to show that the numerator in eq. (F.3) has a zero of order 2 at $u=0$, so the u^{-1} pole is compensated and the integral is always convergent. Although this result guarantees the consistency of eq. (F.3), evaluating the corresponding integral is not trivial. Operating in eq. (F.3), it yields:

$$\begin{aligned}\int_0^q \frac{\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)-1}{u}du &= \int_0^q \frac{\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)}{u}du - \int_0^q \frac{du}{u} \\ &= \frac{1}{2}\int_0^{4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu)} \frac{\exp(-u)}{u}du - \int_0^q \frac{du}{u}, \qquad \text{(F.5)}\end{aligned}$$

where the change of variable $u' = 4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)$ has been used in the first term. Note that the equality in the previous equation cannot be stated strictly speaking, since none of the resulting integrals is convergent. However, from the previous discussion it follows that eq. (F.5) must hold as an equality between finite real numbers when the limit for $u \to 0$ is taken. A primitive for the second integral is obviously the logarithm function. The first one is related to the definition of the exponential integral $\mathrm{E}_i(4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu))$ [Abr72]:

$$\mathrm{E}_i(x) \triangleq \int_x^\infty \frac{\exp(-u)}{u}du \Rightarrow \int_0^x \frac{\exp(-u)}{u}du = \lim_{x\to 0^+}\mathrm{E}_i(x) - \mathrm{E}_i(x). \qquad \text{(F.6)}$$

Although the exponential integral is defined for any complex number, only its evaluation for real numbers is of interest in this dissertation; Fig. F.1 sketches the behavior of the exponential integral for real–valued arguments. Moreover, since $\widetilde{D}(q_0,\xi,\nu)$ is always positive, the argument of E_i is always greater than 0, and the interesting properties of this function are those corresponding to $\mathbb{R}_+$. In particular, the following equivalence holds:

$$\mathrm{E}_i(x) = -\gamma - \log(x) - \sum_{k=1}^{\infty}\frac{(-x)^k}{k\cdot k!}, \text{ if } x>0, \qquad \text{(F.7)}$$

where $\gamma \simeq 0.5772$ is the Euler–Mascheroni constant. The exponential integral shows a logarithmic singularity for $x=0$, according to the behavior shown in Fig. F.1. Casting the expression in eq. (F.7) into eq. (F.5), the integral can be explicitly

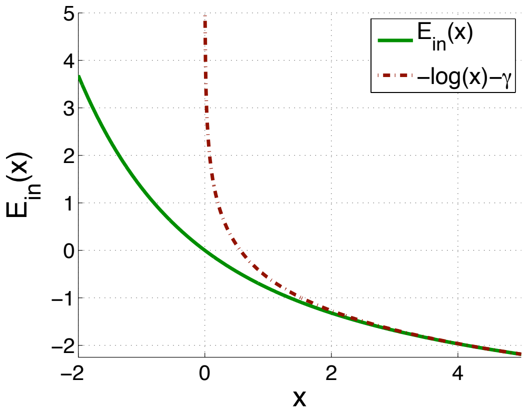


Figure F.2: The non–singular exponential integral $\mathrm{E}_{in}(x)$. For x a large enough, E_{in} rapidly converges to minus the logarithm of x (minus the constant γ).

evaluated:

$$
\begin{aligned}
&\int_0^q \frac{\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)-1}{u}du \\
= &\lim_{q\to 0^+}\left(\frac{1}{2}\mathrm{E}_i(4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu))+\log(q)\right)-\frac{1}{2}\mathrm{E}_i(4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu))-\log(q) \\
= &\lim_{q\to 0^+}\left(\frac{1}{2}\left(-\gamma-\log(4\pi^2\tau\widetilde{D}(q_0,\xi,\nu))-2\log(q)\right.\right. \\
&\qquad\left.\left.-\sum_{k=1}^{\infty}\frac{1}{k\cdot k!}\left(-4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu)\right)^k\right)+\log(q)\right) \\
&-\frac{1}{2}\left(-\gamma-\log(4\pi^2\tau\widetilde{D}(q_0,\xi,\nu))-2\log(q)-\sum_{k=1}^{\infty}\frac{1}{k\cdot k!}\left(-4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu)\right)^k\right)-\log(q) \\
= &\frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k\cdot k!}\left(-4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu)\right)^k.
\end{aligned}
\tag{F.8}
$$

The final series expansion in eq. (F.8) is, by definition, the non–singular exponential integral $\mathrm{E}_{in}(4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu))$:

$$\mathrm{E}_{in}(x)\triangleq\sum_{k=1}^{\infty}\frac{(-x)^k}{k\cdot k!}. \tag{F.9}$$

The non–singular exponential integral is depicted in Fig. F.2; Once again, positive arguments are the space of interest due to the positivity of $\widetilde{D}(q_0,\xi,\nu)$. However, it is very easy to demonstrate that the power series expansion defining E_{in} is absolutely convergent for any real value. It is enough to realize that the absolute value of any term in the series is the corresponding term of the series expansion of the exponential function divided by $k\geq 1$. Each term in the series of E_{in} is minor than the corresponding term in the series of the exponential; since the exponential is absolutely convergent, so it is the non–singular exponential integral.

The final expression in eq. (F.8) could have been obtained directly from eq. (F.4), simply by integrating each term of the power series (this calculation makes sense because the power series normally converges for any $u<\infty$; this property is directly inferred from the normal convergence of the exponential function). However, the representation in terms of the exponential integral allows to derive the asymptotic behavior of $\mathrm{E}_{in}(x)$ for large x in a very easy way. Identifying corresponding terms in eqs. (F.7) and (F.9), the following relation remains evident:

$$\mathrm{E}_{in}(x)=-\gamma-\mathrm{E}_i(x)-\log(x). \tag{F.10}$$

Since the exponential integral $\mathrm{E}_i(x)$ vanishes to 0 very fast as x increases (see Fig. F.1), it is clear that $\mathrm{E}_{in}(x)$ shows a similar behavior to minus the logarithm for relatively high x; in fact, Fig. F.1 suggests that the approximation:

$$\mathrm{E}_{in}(x) \simeq -\gamma - \log(x) \tag{F.11}$$

holds for values of x above 2.

As a concluding remark, the development carried out in this Appendix allows to give a closed–form solution for the integral of eq. (F.1) in terms of the non–singular exponential integral:

$$\begin{aligned}
\widetilde{F_\nu^\Phi}(q,\xi,\nu) &\simeq \frac{-1}{8\pi^2 q}\int_0^q \frac{\Delta_b \exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)}{u}du \\
&= \frac{-1}{8\pi^2 q}\int_0^q \Delta_b \frac{\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)-1}{u}du \\
&= \frac{-1}{8\pi^2 q}\Delta_b \int_0^q \frac{\exp\left(-4\pi^2\tau u^2\widetilde{D}(q_0,\xi,\nu)\right)-1}{u}du \\
&= \frac{-1}{16\pi^2 q}\Delta_b \mathrm{E}_{in}\left(-4\pi^2\tau q^2\widetilde{D}(q_0,\xi,\nu)\right),
\end{aligned} \tag{F.12}$$

which is the final result given in eq. (7.24).

Bibliography

[Abr72] M. Abramowitz and I. A. S. (Eds.). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1972.

[AF07] S. Aja-Fernández, C. Alberola-López, and C.-F. Westin. Signal LMMSE estimation from multiple samples in MRI and DT-MRI. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 4192 of *Lecture Notes in Computer Science*, pp. 368–375. Springer–Verlag, Oct. 2007.

[AF08a] S. Aja-Fernández, C. Alberola-López, and C.-F. Westin. Noise and signal estimation in magnitude MRI and Rician distributed images: A LMMSE approach. *IEEE Transactions on Image Processing*, 17(8):1383–1398, Aug. 2008.

[AF08b] S. Aja-Fernández and K. Krissian. An unbiased Non-Local Means scheme for DWI filtering. In *Proceedings of the Workshop on Computational Diffusion MRI, MICCAI 08*, pp. 277–284, 2008.

[AF08c] S. Aja-Fernández, M. Niethammer, M. Kubicki, M. Shenton, and C.-F. Westin. Restoration of DWI data using a Rician LMMSE estimator. *IEEE Transactions on Medical Imaging*, 27(10):1389–1403, Oct. 2008.

[AF09] S. Aja-Fernández, A. Tristán-Vega, and C. Alberola-López. Noise estimation in single and multiple coil MR data based on statistical models. *Magentic Resonance Imaging*, 2009. To appear.

[Aga09] I. Aganj, C. Lenglet, and G. Sapiro. ODF reconstruction in Q–Ball imaging with solid angle consideration. In *Proceedings of the 6th International Symposium on Biomedical Imaging: from Nano to Macro*, pp. –. IEEE, Jan. 2009. To be published.

[And02] J. Andersson and S. Skare. A model-based method for retrospective correction of geometric distorsions in diffusion-weighted EPI. *NeuroImage*, 16:177–199, 2002.

[And05] A. Anderson. Measurement of fiber orientation distributions using High Angular Resolution Diffusion Imaging. *Magnetic Resonance in Medicine*, 54(5):1194–1206, 2005.

[Bar08] A. Barmpoutis, B. Vemuri, D. Howland, and J. Forder. Extracting tractosemas from a displacement probability field for tractography in DW–MRI. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 5241 of *Lecture Notes in Computer Science*, pp. 9–16. Springer–Verlag, Sep. 2008.

[Bas94a] P. Basser, J. Mattiello, and D. Lebihan. MR diffusion tensor spectroscopy and imaging. *Biophysial Journal*, 66(1):259–267, Jan. 1994.

[Bas94b] P. J. Basser, J. Mattiello, and D. L. Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance*, 103(3):247–254, 1994.

[Bas96] P. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by Quantitative–Diffusion–Tensor MRI. *Journal of Magnetic Resonance*, 111(3):209–219, Jun. 1996.

[Bas00a] P. Basser and S. Pajevic. Statistical artifacts in diffusion tensor MRI (DT–MRI) caused by background noise. *Magnetic Resonance in Medicine*, 44:41–50, 2000.

[Bas00b] P. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT–MRI data. *Magnetic Resonance in Medicine*, 44(4):625–632, Oct. 2000.

[Bas06] S. Basu, T. Fletcher, and R. Whitaker. Rician noise removal in diffusion tensor MRI. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 1 of *Lecture Notes in Computer Science*, pp. 117–125. Springer–Verlag, 2006.

[Ber07a] Ø. Bergmann, G. Kindlmann, A. Lundervold, and C.-F. Westin. Diffusion k-tensor estimation from Q–ball imaging using discretized principal axes. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 4191 of *Lecture Notes in Computer Science*, pp. 268–275. Springer–Verlag, Oct. 2007.

[Ber07b] Ø. Bergmann, G. Kindlmann, S. Peled, and C.-F. Westin. Two–tensor fiber tractography. In *Proceedings of the 4th International Symposium on Biomedical Imaging: from Nano to Macro*, pp. 796–799. IEEE, Apr. 2007.

[Ber08] J. Berman, S. chung, P. Mukherjee, C. Hess, E. Han, and R. Henry. Probabilistic streamline q–ball tractography using the residual bootstrap. *NeuroImage*, 39:215–222, 2008.

[Bla04] M. Blaimer, F. Breuer, M. Mueller, R. Heidemann, M. Griswold, and P. Jakob. SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method. *Top Magnetic Resonance Imaging*, 15(4):223–236, Aug. 2004.

[Blo46] F. Bloch. Nuclear induction. *Physical Review*, 70:460–474, 1946.

[Blo08] L. Bloy and R. Verma. On computing the underlying fiber directions from the diffusion orientation distribution function. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 5241 of *Lecture Notes in Computer Science*, pp. 1–8. Springer–Verlag, Sep. 2008.

[Bro04] R. Brown, Y. Wang, P. Spincemaille, and R. Lee. On the noise correlation matrix for multiple radio frequency coils. *Magnetic Resonance in Medicine*, 58:218–224, 2004.

[Bru93] M. Brummer, R. Mersereau, R. Eisner, and R. Lewine. Automatic detection of brain contours in MRI data sets. *IEEE Transactions on Medical Imaging*, 12(2):153–166, 1993.

[Bua05] A. Buades, B. Coll, and J. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation*, 4(2):490–530, 2005.

[Buc98] M. Buchsbaum, C. Tang, S. Peled, H. Gudbjartsson, D. Lu, E. H. J. Downhill, M. Haznedar, J. Fallon, and S. Atlas. MRI white matter diffusion anisotropy and PET metabolic rate in schizophrenia. *Neuroreport*, 9(3):425–430, 1998.

[Byd02] M. Bydder, D. Larkman, and J. Hajnal. Generalized SMASH imaging. *Magnetic Resonance in Medicine*, 47:160–170, 2002.

[CA08] J. Cohen-Adad, M. Descoteaux, S. Rossignol, R. Hoge, R. Deriche, and H. Benali. Detection of multiple pathways in the spinal cord using q–ball imaging. *NeuroImage*, 42:739–749, 2008.

[Cal91] P. Callahan. *Principles of Nuclear Magnetic Resonance Microscopy*. Clarendon Press, Oxford, 1991.

[Che02] C. Chefd'hotel, D. Tschumperle, R. Deriche, and O. Faugueras. Constrained flows of matrix–valued functions: application to diffusion tensor regularization. In *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, vol. 2350 of *Lecture Notes in Computer Science*, pp. 251–265. Springer–Verlag, 2002.

[Che05] B. Chen and E. Hsu. Noise removal in magnetic resonance diffusion tensor imaging. *Magnetic Resonance in Medicine*, 54:393–407, 2005.

[Cla08] R. Clarke, P. Scifo, G. Rizzo, F. Dell'Acqua, G. Scotti, and F. Fazio. Noise correction on Rician distributed data for fibre orientation estimators. *IEEE Transactions on Medical Imaging*, 27(9):1242–1251, Sep. 2008.

[CM07] C. Castaño-Moraga, C. Lenglet, R. Deriche, and J. Ruiz-Alzola. A Riemannian approach to anisotropic filtering of tensor fields. *Signal Processing*, 87(2):263–276, 2007.

[Col98] D. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes, and A. Evans. Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging*, 17(3):463–468, 1998.

[Con97] C. Constantinides, E. Atalar, and E. McVeigh. Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magnetic Resonance in Medicine*, 38:852–857, 1997.

[Cou04] O. Coulon, D. Alexander, and S. Arridge. Diffusion tensor magnetic resonance image regularization. *Medical Image Analysis*, 8(1):47–67, 2004.

[Cou08] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. An optimized blockwise non local means denoising filter for 3D magnetic resonance images. *IEEE Transactions on Medical Imaging*, 27(4):425–441, Apr. 2008.

[CR09] E. Canales-Rodríguez, L. Melie-García, and Y. Iturria-Medina. Mathematical description of q-space in spherical coordinates: exact Q-Ball imaging. *Magnetic Resonance in Medicine*, 61:1350–1367, 2009.

[Des06] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. Apparent Diffusion Profile estimation from High Angular Resolution Diffusion Images: estimation and applications. *Magnetic Resonance in Medicine*, 56(2):395–410, 2006.

[Des07] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. Regularized, fast, and robust analytical Q-Ball imaging. *Magnetic Resonance in Medicine*, 58:497–510, 2007.

[Des08] M. Descotaux, N. Wiest-Daesslé, S. Prima, C. Barillot, and R. Deriche. Impact of Rician adapted Non–Local Means filtering on HARDI. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 5242 of *Lecure Notes in Computer Science*, pp. 122–130. Springer–Verlag, 2008.

[Des09] M. Descoteaux, R. Deriche, T. Knösche, and A. Anwander. Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE Transactions on Medical Imaging*, 28(2):269–286, Feb. 2009.

[Die08] O. Dietrich, J. Raya, S. Reeder, M. Ingrisch, M. Reiser, and S. Schoenberg. Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics. *Magnetic Resonance Imaging*, 26:754–762, 2008.

[Dru93] D. Drumheller. General expressions for Rician density and distribution functions. *IEEE Transactions on Aerospace and Electronic Systems*, 29(2):580–588, Apr. 1993.

[Fil07] P. Fillard, X. Pennec, V. Arsigny, and N. Ayache. Clinical DT–MRI estimation, smoothing, and fiber tracking with log–euclidean metrics. *IEEE Transactions on Medical Imaging*, 26(11):1472–1482, Nov. 2007.

[Fra02] L. Frank. Characterization of anisotropy in High Angular Resolution Diffusion–weighted MRI. *Magnetic Resonance in Medicine*, 47(6):1083–1099, Jan. 2002.

[Fri06] O. Friman, G. Farneback, and C.-F. Westin. A Bayesian approach to stochastic white matter tractography. *IEEE Transactions on Medical Imaging*, 25(8):965–978, Aug. 2006.

[Ger92] G. Gerig, O. Kübler, R. Kikinis, and F. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, Jun. 1992.

[Gil07] G. Gilbert, D. Simard, and G. Beaudoin. Impact of an improved combination of signal from array coils in diffusion tensor imaging. *IEEE Transactions on Medical Imaging*, 26(11):1428–1436, Nov. 2007.

[Gri02] M. Griswold, P. Jakob, R. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalize autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47:1202–1210, 2002.

[Gud95] H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34:910–914, 1995.

[Han50] R. Hanh. Spin echos. *Physical Review*, 80:545–554, 1950.

[Har92] M. Harpen. Noise correlations exist for independent RF coils. *Magnetic Resonance in Medicine*, 23:394–397, 1992.

[Hay90] C. Hayes and P. Roemer. Noise correlations in data simultaneously acquired from multiple surface coil arrays. *Magnetic Resonance in Medicine*, 16:181–191, 1990.

[Hog05] W. Hoge, D. Brooks, B. Madore, and W. Kyriakos. A tour of accelerated parallel MR imaging from a linear systems perspective. *Concepts of Magnetic Resonance*, 27A(1):17–37, Sep. 2005.

[Hor02] M. Horsfiled and D. Jones. Applications of diffusion-weighted and diffusion tensor MRI to white matter diseases - a review. *Nuclear Magnetic Resonance in Biomedicine*, 15:570–577, 2002.

[Iba05] L. Ibañez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide*. Kitware, Inc. ISBN 1-930934-15-7, http://www.itk.org/ItkSoftwareGuide.pdf, second ed., 2005.

[Jan03] K. Jansons and D. Alexander. Persistent Angular Structures: new insights from diffusion magnetic resonance imaging data. *Inverse Problems*, 19:1031–1046, 2003.

[Ji07] J. Ji, B. Son, and S. Rane. PULSAR: a MatLab toolbox for parallel magnetic resonance imaging using array coils and multiple channel receivers. *Concepts in Magnetic Resonance Part B: Magnetic Resonance Engineering*, 1:24–36, 2007.

[Jia07] B. Jian, B. Vemuri, E. Özarslan, P. Carney, and T. Marecy. A novel tensor distribution model for the diffusion–weighted MR signal. *NeuroImage*, 37(1):164–176, Aug. 2007.

[Jon04] D. Jones and P. Basser. Squashing peanuts and smashing pumpkins: how noise distorts diffusion weighted MR data. *Magnetic Resonance in Medicine*, 52:979–993, 2004.

[Jon05] L. Jonasson. *Segmentation of Diffusion Weighted MRI using the level set framework*. Ph.D. thesis, École Polytechnique Fédérale de Lausanne, Laussane, Swiss, 2005.

[Kay93] S. Kay. *Fundamentals of Statistical Signal Processing. Estimation Theory.* Prentice–Hall, Upper Saddle River, New Jersey 07458 (USA), 1993.

[Kin84] K. King and P. Moran. A unified description of NMR imaging, data collection strategies and reconstruction. *Medical Physics*, 11(1):1–14, 1984.

[Kre05] B. Kreher, J. Schneider, I. Mader, E. Martin, J. Henning, and K. Il'yasov. Multitensor approach for analysis and tracking of complex fiber configurations. *Magnetic Resonance in Medicine*, 54(5):1216–1225, Sep. 2005.

[Kri09] K. Krissian and S. Aja-Fernández. Noise-driven anisotropic diffusion filtering of MRI. *IEEE Transactions on Image Processing*, 2009. To Appear. DOI: 10.1109/TIP.2009.2025553.

[Kub07] M. Kubicki, R. McCarley, C.-F. Westin, H.-J. Park, S. Maier, R. Kikinis, F. Jolesz, and M. Shenton. A review of diffusion tensor imaging studies in schizophrenia. *Journal of Psychiatric Research*, 41(1–2):15–30, 2007.

[Lan07] B. Landman, P.-L. Bazin, J. Prince, and J. Hopkins. Diffusion tensor estimation by maximizing Rician likelihood. In *Procs. of IEEE 11th International Conference on Computer Vision*, Oct. 2007.

[Lar07] D. Larkman and R. Nunes. Parallel magnetic resonance imaging. *Physics of Medical Biology*, 52:15–55, 2007. Invited topical review.

[Lau73] P. Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.

[LeB86] D. LeBihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407, Nov. 1986.

[Lia00] Z.-P. Liang and P. Lauterbur. *Principles of Magnetic Resonance Imaging.* IEEE Press, Ney York, USA, 2000.

[Lim02] K. Lim and J. Helpern. Neuropsychiatric applications of DTI – a review. *Nuclear Magnetic Resonance in Biomedicine*, 15(7–8):587–593, 2002.

[Man73] P. Mansfield and P. Grannel. NMR diffraction in solids. *Journal of Physical Chemics*, 6:422–426, 1973.

[Man08] J. Manjón, J. Carbonell-Caballero, J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles. MRI denoising using Non-Local Means. *Medical Image Analysis*, 12:514–523, 2008.

[Mar95] T. Marzetta. EM algorithm for estimating the parameters of multivariate complex Rician density for polarimetric SAR. In *Proceedings of ICASSP*, vol. 5, pp. 3651–3654, 1995.

[MC08] S. Merino-Caviedes and M. Martín-Fernández. *Tensors in Image Processing and Computer Vision*, chap. User interfaces to interact with tensor fields, pp. 429–454. Springer Verlag, London, 2008.

[McG93] G. McGibney and M. Smith. Unbiased signal to noise ratio measure for magentic resonance images. *Medical Physics*, 20(4):1077–1078, Jul. 1993.

[MF04] M. Martín-Fernández, C.-F. Westin, and C. Alberola-López. 3D Bayesian regularization of diffusion tensor MRI using multivariate Gaussian Markov random fields. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 3216 of *Lecture Notes in Computer Science*, pp. 351–359. Springer–Verlag, 2004.

[MF07] M. Martín-Fernández, C. Alberola-López, J. Ruiz-Alzola, and C.-F. Westin. Sequential anisotropic Wiener filtering applied to 3D MRI data. *Magnetic Resonance in Medicine*, 25:278–292, 2007.

[MF09] M. Martín-Fernández, E. Muñoz-Moreno, L. Cammoun, J.-P. Thiran, C.-F. Westin, and C. Alberola-López. Sequential anisotropic multichannel Wiener filtering with Rician bias correction applied to 3D regularization of DWI data. *Medical Image Analysis*, 13:19–35, 2009.

[Mor02] S. Mori and P. van Zijl. Fiber tracking: principles and strategies – a technical review. *Nuclear Magnetic Resonance in Biomedicine*, 15:468–480, 2002.

[Mor05] S. Mori, S. Wakana, L. Nagae-Poetscher, and P. van Zijl. *MRI Atlas of Human White Matter*. Elsevier, Amsterdam, The Netherlands, 2005.

[Net04] T. Netsch and A. van Muiswinkel. Quantitative evaluation of image-based distortion correction in Diffusion Tensor Imaging. *IEEE Transactions on Medical Imaging*, 23(7):789–798, Jul. 2004.

[NG08] S. Nedjani-Gilani, D. Alexander, and G. Parker. Regularized super–resolution for diffusion MRI. In *Proceedings of the 5th International Symposium on Biomedical Imaging: from Nano to Macro*, pp. 875–878. IEEE, May 2008.

[Nie96] T. Niendorf, R. Dijkhuizen, D. Norris, M. van Lookeren Campagne, and K. Nicolay. Biexponential diffusion attenuation in various states of brain tissue: implications for diffusion–weighted imaging. *Magnetic Resonance in Medicine*, 36(6):847–857, 1996.

[Nie04] J. Nielsen, N. Ghugre, and A. Panigrahy. Affine and polynomial mutual information coregistration for artifact elimination in diffusion tensor imaging of newborns. *Magnetic Resonance Imaging*, 22:1319–1323, Aug. 2004.

[Now99] R. Nowak. Wavelet-based Rician noise removal for Magnetic Resonance Imaging. *IEEE Transactions on Image Processing*, 8(10):108–1419, Oct. 1999.

[Ots79] N. Otsu. A threshold selection method from gray–level histogram. *IEEE Transactions on Systems, Man, and Cybernetics*, 9:62–66, 1979.

[Öza05] E. Özarslan, B. Vemuri, and T. Mareci. Generalized scalar measures for diffusion MRI using trace, variance, and entropy. *Journal of Magnetic Resonance in Medicine*, 53:866–876, 2005.

[Öza06] E. Özarslan, T. Sepherd, B. Vemuri, S. Blackband, and T. Mareci. Resolution of complex tissue microarchitecture using the Diffusion Orientation Transform (DOT). *NeuroImage*, 31:1086–1103, 2006.

[Pag05] E. Pagani, M. Filippi, M. Rocca, and M. Horsfiled. A method for obtainig tract-specific diffusion tensor MRI measurements in the presence of disease: applications to patients with clinically isolated syndromes suggestive of multiple sclerosis. *NeuroImage*, 26(1):258–265, 2005.

[Pap91] A. Papoulis. *Probability, random variables, and stochastic processes.* MacGraw–Hill, New York, NY, USA, third ed., 1991.

[Pel06] S. Peled, O. Friman, F. Jolesz, and C.-F. Westin. Geometrically constrained two–tensor model for crossing tracts in DWI. *Magnetic Resonance in Medicine*, 24(9):1263–1270, 2006.

[Pen06] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66:41–66, 2006.

[Piž03] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy. A versatile Wavelet domain noise filtration technique for medical imaging. *IEEE Transactions on Medical Imaging*, 22(3):323–331, Mar. 2003.

[Pos95] S. Posse, G. Tedeschi, R. Risinger, R. Ogg, and D. LeBihan. High speed ^{1}H spectroscopic imaging in human brain by echo planar spatial-spectral encoding. *Magnetic Resonance in Medicine*, 33:34–40, 1995.

[Pou00] C. Poupon, C. Clark, V. Frouin, J. Rgis, I. Bloch, D. L. Bihan, and J.-F. Mangin. Regularization of diffusion–based direction maps for the tracking of brain white matter fascicles. *NeuroImage*, 12(2):184–195, 2000.

[Prč08] V. Prčkovska, A. Roebroeck, W. Pullens, A. Vilanova, and B. ter Haar Romeny. Optimal acquisition schemes in High Angular Resolution Diffusion Weighted Imaging. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 5242 of *Lecture Notes in Computer Science*, pp. 9–17. Springer–Verlag, Sep. 2008.

[Pru99] K. Pruessmann, M. Weiger, M. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42:952–962, 1999.

[Pru01] K. Pruessmann, M. Weiger, P. Bornert, and P. Boesiger. Advances in sensitivity encoding with arbitrary k–space trajectories. *Magnetic Resonance in Medicine*, 46:638–651, 2001.

[Pur46] E. Purcell, H. Torrey, and R. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical Review*, 69:37, 1946.

[Red92] T. Redpath. Noise correlation in multicoil receiver systems. *Magnetic Resonance in Medicine*, 24:85–89, 1992.

[Roe90] P. Roemer, W. Edelstein, C. Hayes, S. Souza, and O. Mueller. The NMR phased array. *Magnetic Resonance in Medicine*, 16:192–225, 1990.

[Sal05] R. Salvador, A. Peña, D.-K. Menon, T.-A. Carpenter, J.-D. Pickard, and E.-T. Bullmore. Formal characterization and extension of the linearized diffusion tensor model. *Human Brain Mapping*, 24:144–155, 2005.

[Shi09] H. Shinmoto, K. Oshio, A. Tanimoto, N. Higuchi, S. Okuda, S. Kuribayashi, and R. Mulkern. Biexponential apparent diffusion coefficients in prostate cancer. *Magnetic Resonance Imaging*, 27:355–359, 2009.

[Sij98] J. Sijbers, A.-J. den Dekker, J. Van Audekerke, M. Verhoye, and D. Van Dyck. Estimation of the noise in magnitude MR images. *Magentic Resonance Imaging*, 16(1):87–90, 1998.

[Sij04] J. Sijbers and A.-J. den Dekker. Maximum Likelihood estimation of signal amplitude and noise variance from MR data. *Magnetic Resonance Imaging*, 51:586–594, 2004.

[Sij07] J. Sijbers, D. Poot, A.-J. den Dekker, and W. Pintjenst. Automatic estimation of the poise variance from the histogram of a magnetic resonance image. *Physics in Medicine and Biology*, 52:1335–1348, Feb. 2007.

[Smi06] S. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. Nichols, C. Mackay, K. Watkins, O. Cicarelli, M. Cader, P. Matthews, and T. Behrens. Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31(4):1487–1505, 2006.

[Smi07] S. Smith, H. Johansen-Berg, M. Jenkinson, D. Rueckert, T. Nichols, K. Miller, M. Robson, D. Jones, J. Klein, A. Bartsch, and T. Behrens. Acquisition and voxelwise analysis of multi-subject diffusion data with tract based spatial statistics. *Nature Protocols*, 2(3):499–503, 2007.

[Söd95] O. Söderman and B. Jönsson. Restricted diffusion in cylindrical geometry. *Journal of Magnetic Resonance*, 117:94–97, 1995.

[Sod97] D. Sodickson and W. Manning. Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38:591–603, 1997.

[Ste65] E.-O. Stejskal and J.-E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *Journal of Chemical Physics*, 42:288–292, 1965.

[Ste74] M. Stephens. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974.

[Sun04] P. Sundgren, Q. Dong, D. Gómez-Hassan, S. Mukherji, P. Maly, and R. Welsh. Diffusion tensor imaging of the brain: review of clinical applications. *Neuroradiology*, 46(5):339–350, 2004.

[Tay04] W. Taylo, E. Hsu, K. Krishnan, and J. MacFall. Diffusion tensor imaging: background, potential and utility in psychiatric research. *Biological Psychiatry*, 55(3):201–207, 2004.

[Ten02] C. Tench, P. Morgan, M. Wilson, and L. Blumhardt. White matter mapping using diffusion tensor MRI. *Magnetic Resonance in Medicine*, 47:967–972, 2002.

[Thu07] P. Thunberg and P. Zetterberg. Noise distribution in SENSE– and GRAPPA–reconstructed images: a computer simulation study. *Magnetic Resonance Imaging*, 25:1089–1094, 2007.

[Tou07] J.-D. Tournier, F. Calamante, and A. Connelly. Robust determination of the fibre orientation distribution in diffusion MRI: Non–negativity constrained super–resolved spherical deconvolution. *NeuroImage*, 35:1459–1472, 2007.

[Tou08] J.-D. Tournier, C.-H. Yeh, F. Calamante, K.-H. Cho, A. Connelly, and C.-P. Lin. Resolving crossing fibres using constrained spherical deconvolution: Validation using diffusion–weighted imaging phantom data. *NeuroImage*, 42:617–625, 2008.

[Tuc03] D. Tuch, T. Reese, M. Wiegell, and V. Wedeen. Diffusion MRI of complex neural architecture. *Neuron*, 40:885–895, 2003.

[Tuc04] D. Tuch. Q–Ball imaging. *Magnetic Resonance in Medicine*, 52:1358–1372, 2004.

[TV08] A. Tristán-Vega and S. Aja-Fernández. Joint LMMSE estimation of DWI data for DTI processing. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 5242 of *Lecture Notes in Computer Science*, pp. 27–34. Springer–Verlag, Sep. 2008.

[TV09a] A. Tristán-Vega and S. Aja-Fernández. Design and construction of a realistic DWI phantom for filtering performance assessment. In *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science. Springer–Verlag, Sep. 2009. To appear.

[TV09b] A. Tristán-Vega, S. Aja-Fernández, and C.-F. Westin. On the blurring of the Funk–Radon transform in Q–ball imaging. In *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science. Springer–Verlag, Sep. 2009. To appear.

[TV09c] A. Tristán-Vega, C.-F. Westin, and S. Aja-Fernández. Bias of least squares approaches for diffusion tensor estimation from array coils in DT–MRI. In *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science. Springer–Verlag, Sep. 2009. To appear.

[TV09d] A. Tristán-Vega, C.-F. Westin, and S. Aja-Fernández. Estimation of fiber orientation probability density functions in High Angular Resolution Diffusion Imaging. *NeuroImage*, 47:638–650, 2009.

[Wan00] Y. Wang. Description of parallel imaging in MRI using multiple coils. *Magnetic Resonance in Medicine*, 44:495–499, 2000.

[Wan01] J. Wang, T. Kluge, M. Nittka, V. Jellus, B. Kühn, and B. Kiefer. Parallel acquisition techniques with modified SENSE reconstructions: mSENSE. In *Proceedings of First Wurzburg Workshop*, p. 92, 2001.

[Wan04a] Z. Wang, A.-C. Bovik, H.-R. Sheikh, and E.-P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[Wan04b] Z. Wang, B. Vemuri, Y. Chen, and T. Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *IEEE Transactions on Medical Imaging*, 23(8):930–939, 2004.

[WD07] N. Wiest-Daesslé, S. Prima, P. Coupé, S. Morrissey, and C. Barillot. Non-local means variants for denoising of diffusion-weighted and diffusion tensor MRI. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 4792 of *Lecture Notes in Computer Science*, pp. 344–351. Springer–Verlag, 2007.

[WD08] N. Wiest-Daesslé, S. Prima, P. Coupé, S. Morrissey, and C. Barillot. Rician noise removal by Non–Local Means filtering for low signal–to–noise ratio MRI: Applications to DT–MRI. In *Medical Image Computing and Computer-Assisted Intervention*, vol. 5242 of *Lecture Notes in Computer Science*, pp. 171–179. Springer–Verlag, 2008.

[Wed05] V. Wedeen, P. Hagmann, W.-Y. Tseng, T. Reese, and R. Weisskoff. Mapping complex tissue architecture with diffusion spectrum imaging. *Magnetic Resonance in Medicine*, 54:1377–1386, 2005.

[Wes01] C.-F. Westin, S. Maier, H. Mamata, A. Nabavi, F. Jolesz, and R. Kikinis. Processing and visualization for diffusion tensor MRI. *Medical Image Analysis*, 6:93–108, 2001.

[Xu08] H. Xu, J. Xu, and F. Wu. On the biased estimation of Nonlocal Means filter. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1149–1152, 2008.

[Zha09] F. Zhang, E. Hancock, C. Goodlett, and G. Gerig. Probabilistic white matter fiber tracking using particle filtering and Von Mishes–Fisher sampling. *Medical Image Analysis*, 13:5–18, 2009.