# HARDI Denoising using Non-local Means on the $\mathbb{R}^3 \times \mathbb{S}^2$ Manifold

by

Alan Kuurstra

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

**Abstract**

Magnetic resonance imaging (MRI) has long become one of the most powerful and accurate tools of medical diagnostic imaging. Central to the diagnostic capabilities of MRI is the notion of contrast, which is determined by the biochemical composition of examined tissue as well as by its morphology. Despite the importance of the prevalent $T_1$, $T_2$, and proton density contrast mechanisms to clinical diagnosis, none of them has demonstrated effectiveness in delineating the morphological structure of the white matter – the information which is known to be related to a wide spectrum of brain-related disorders. It is only with the recent advent of diffusion-weighted MRI that scientists have been able to perform quantitative measurements of the diffusivity of white matter, making possible the structural delineation of neural fibre tracts in the human brain. One diffusion imaging technique in particular, namely high angular resolution diffusion imaging (HARDI), has inspired a substantial number of processing methods capable of obtaining the orientational information of multiple fibres within a single voxel while boasting minimal acquisition requirements.

HARDI characterization of fibre morphology can be enhanced by increasing spatial and angular resolutions, however, doing so drastically reduces the signal-to-noise ratio. Since pronounced measurement noise tends to obscure and distort diagnostically relevant details of diffusion-weighted MR signals, increasing spatial or angular resolution necessitates application of the efficient and reliable tools of image denoising. The aim of this work is to develop an effective framework for the filtering of HARDI measurement noise which takes into account both the manifold to which the HARDI signal belongs and the statistical nature of MRI noise. These goals are accomplished using an approach rooted in non-local means (NLM) weighted averaging. The average includes samples, and therefore dependencies, from the entire manifold and the result of the average is used to deduce an estimate of the original signal value in accordance with MRI statistics. NLM averaging weights are determined adaptively based on a neighbourhood similarity measure. The novel neighbourhood comparison proposed in this thesis is one of spherical neighbourhoods, which assigns large weights to samples with similar local orientational diffusion characteristics. Moreover, the weights are designed to be invariant to both spatial rotations as well as to the particular sampling scheme in use. This thesis provides a detailed description of the proposed filtering procedure as well as experimental results with synthetic and real-life data. It is demonstrated that the proposed filter has substantially better denoising capabilities as compared to a number of alternative methods.

**Acknowledgements**

I would like to express my gratitude towards Dr. Oleg Michailovich, who has provided his continuous support and insight into my research interests and activities. Over the past two years, his technical expertise has been instrumental to my growth as an aspiring researcher. Thanks to my fellow graduate students and friends: Sudipto Dolui, James Ho, Tariq Nanji, Igor Solovey, and Jaimal Soni. More important than the fruitful discussions regarding research, they have been like family and I have truly enjoyed journeying through the graduate experience with them. Thanks to my friends back home. The road trips, cottage weekends, themed parties, and other shenanigans have somehow managed to keep me sane. Finally, I would like to thank my family, John, Heather, Philip, Natalie, and Michael Kuurstra. Without their love and support, the work presented here would not have been possible.

***To my family, who have shared my tears of both laughter and sadness,***
***for never failing to provide me with their love, support, and encouragement.***

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Magnetic resonance imaging (MRI) has long become one of the most powerful and accurate tools of medical diagnostic imaging. As MRI employs neither ionizing radiation nor excessive levels of electromagnetic energy, it is known to be practically harmless to the human body which makes it applicable to diverse groups of patients including infants, high-risk pregnant women, and elders. Furthermore, recent advances in the design of magnetic coils, along with the advent of sophisticated tomographic techniques, has allowed MRI to achieve sub-millimetre resolution, which has made it an extremely powerful diagnostic tool for early detection of abnormal developments in the brain. The spectrum of neurological disorders which can be currently diagnosed based on the information provided by MRI is impressively broad [52, 67, 68, 86, 96]. Central to the diagnostic capabilities of MRI is the notion of *contrast*, which is determined by the biochemical composition of examined tissue as well as by its morphology. Prevalent in the current MRI practice are the contrasts determined by $T_1/T_2$ relaxation times and proton density (PD). Despite the exceptional importance of these contrast mechanisms to clinical diagnosis, none of them has demonstrated effectiveness in delineating the morphological structure of the white matter – the information which is known to be related to a wide spectrum of brain-related disorders [7, 36, 38, 69, 78, 84, 105, 106]. It is only with the recent advent of *diffusion-weighted MRI* (DW-MRI) that scientists have been able to perform quantitative measurements of the diffusivity of white matter, making possible the structural delineation of neural fibre tracts in the human brain. [9, 10, 14].

The human brain consists of about $10^{11}$ nerve cells that can be subdivided into about 1000 different cell types, a complexity that far exceeds any other organ in the body. A further complexity is evident in the way in which neuron cells of the brain interconnect and function [51]. In contrast to other types of cells, each neuron communicates with its target

cells by means of a protoplasmic protrusion, called an axon. Moreover, axons with similar destinations tend to form bundles, known as *neural fibre tracts*. Thus, for example, the neural bundles connecting two cortical regions are called U-fibres (between adjacent gyri), association fibres (between different lobes), or commissural fibres (between the right and left hemispheres), while those connecting cortex and deep-brain regions (e.g., the cortex and thalamus or cortex and spinal cord) are called projection fibres [102]. Taken together, these fibres play a fundamental role in the determination of brain connectivity, which is still far from being completely understood. Yet, it is nowadays believed that through reconstructing the pattern of connectivity of the neural fibre tracts in both healthy and diseased subjects, it is possible to obtain an abundance of valuable diagnostic information which can be used for early diagnostics of brain-related disorders, for assessing the damage caused to the brain by stroke, tumours or injuries, as well as for planning and monitoring of neurosurgeries [52, 67].

Neural fibre tracts form a principal part of the white matter, whose structural delineation is critically important for our understanding of brain connectivity. Unfortunately, traditional MRI is incapable of revealing the detailed anatomy of the white matter, since the $T_1$,$T_2$ and PD contrasts used in traditional MRI are only sensitive to the chemical composition of cerebral tissue, while the white matter is chemically uniform [52, Ch.5]. DW-MRI, on the other hand, can generate contrast based on the phenomenon of water diffusion, which is an essential physical process for the normal functioning of living systems. In particular, DW-MRI is capable of measuring the *apparent diffusion coefficients* (ADC) of cerebral tissue, which depend on the local tissue microstructure and can therefore be used as a probe of the physical properties of the white matter. Furthermore, since water molecules tend to diffuse more freely along the direction of neural fibres, if one can quantify the orientational preferences of diffusion, it is possible to deduce the axonal orientations [67].

Establishing the main principles of diffusion imaging dates back to 1965, when Dr. Stejskal and Dr. Tanner proposed to measure the diffusion constant of water molecules, commonly referred to as spins, using nuclear magnetic resonance and a pulsed magnetic field gradient system [83]. Their measurement experiment – known as Pulsed Gradient Spin Echo (PGSE) – is still the basis of diffusion protocols exploited by modern MRI scanners. A principal attribute of PGSE is the application of a pair of *diffusion-encoding* magnetic gradients. Whereas the first gradient is used to encode the spatial positions of the spins, the second gradient tries to reverse the effect of spatial encoding after the spins are allowed to diffuse for a predetermined amount of time known as the *mixing* time. The measurement of spin displacement is manifested in an attenuation of the MR readout known as the *diffusion weighting*. Diffusion weighting depends both on the amount of spin diffusion and

on parameters related to the diffusion-encoding gradients. The latter is captured in the ***q**-value* and determines a measurement's location in the ***q**-space*. One of the most significant results of the theory of diffusion imaging is that if one can measure the DW-MR signal over a sufficiently large set of points in **q**-space, it is then possible to recover the *ensemble average (diffusion) propagator* (EAP) – the function which quantifies the probability of a spin being found at position $\mathbf{r}\, d\mathbf{r}$ after the mixing time [83]. In particular, the DW-MR signal is related to the EAP via Fourier transformation [25]. Needless to say, since the EAP depends on the local microstructure of interrogated tissue, it is essential for understanding the tissue composition and organization.

*Diffusion spectral imaging* (DSI) [95] attempts to recover the EAP directly, while making no assumptions about tissue microstructure or the functional shape of the EAP. The recovery can be performed by sampling the attenuation signal over a Cartesian grid in the **q**-space, followed by the application of fast Fourier transformation. An alternative way to sample the **q**-space is to put the sampling points on the surfaces of concentric spherical shells [33]. It goes without saying that the major limitation of DSI is its acquisition requirements, demanding roughly 500 DW-images with each acquisition's diffusion-encoding gradient corresponding to a point in **q**-space. Unfortunately, such acquisition requirements put DSI far beyond the scope of practical applications of DW-MRI. It is clearly a very unfortunate fact, since DSI is the only method of DW-MRI allowing a model-free reconstruction of the EAP, which is a fundamental descriptor of cerebral diffusion.

In *diffusion tensor imaging* (DTI) [2, 9, 14, 15, 53, 66, 73, 98], the EAP is assumed to be a Gaussian probability density. With this assumption, the DW-MR signal is completely described by the related $3 \times 3$ *diffusion tensor*. Recovering the diffusion tensor requires six diffusion-encoding gradients and it is typical to restrict the corresponding **q**-values to belong to a single spherical shell in **q**-space. DTI is capable of capturing the orientation of neural fibre tracts, and hence the neuroanatomy of the white matter. Moreover, the values of the diffusion tensor can provide information on the spatial pattern of diffusion flow, which can be described in terms of relative and fractional anisotropy, lattice index, as well as by linear, spherical, and planar diffusion measures [15,97]. What is even more important is that all the above measures can be computed along specific fibre tracts based on the results of *fibre tractography* [11, 13, 59–61, 81]. The spectrum of neurological disorders which are now much better understood due to DTI is impressively broad. Alzheimer's dementia and non-Alzheimer forms of degenerative dementia [69,78], multiple sclerosis [38], amyotrophic lateral sclerosis [36], ischemic stroke [2], epilepsy [7], brain tumours [67], movement disorders [105], and schizophrenia [76] are only a few examples in a long list of such diseases (for a thorough overview see, e.g., [52, 84]). Unfortunately, since the DTI model assumes water diffusion to be unimodal, its application at the spatial locations

where neural fibres cross, touch upon each other, or diverge is known to result in severe underestimation artifacts [3, 4, 39, 40, 92, 93, 98].

To resolve the shortcomings of DTI, the method of *High Angular Resolution Diffusion Imaging* (HARDI) has been proposed [6, 31, 40, 92, 93]. At the practical level, HARDI is similar to DTI except that it requires a considerably higher number of encoding gradients, typically between 80 and 100. The availability of more samples in the **q**-space allows for the use of more accurate and realistic models to describe the EAP. Specifically, HARDI data can be used to approximate the DW-MR attenuation signal as a function of spatial orientation $\mathbf{u} \in \mathbb{S}^{\mathbf{2}}$ and, subsequently, subject the resulting estimate to the Funk-Radon transform. The significance of this computation is in the fact that, with a very high accuracy, the result approximates the *orientation distribution function* (ODF) of diffusion flow, which is related to the EAP through a radial projection onto the spatial direction defined by **u**. Thus, while the EAP quantifies the likelihood with which a spin is displaced by **r**, the ODF quantifies the likelihood with which the spin is expected to move in direction **u**. Consequently, the local maxima of the ODF can be expected to point at the directions of the most likely diffusion flows. Despite the fact that the ODF is only a projection of the true EAP, it is still capable of resolving multiple diffusion flows (and, hence, fibre orientations) within a single voxel – an analysis which is impossible to achieve by means of DTI. For this reason, the advent of HARDI has revolutionized neural fibre tractography, whose precision has since been improved by orders of magnitude [59–61].

Although the last two decades has seen DW-MRI develop into an established research tool with a great impact on health care and neuroscience, like any other MRI technique it remains subject to artifacts and pitfalls [16, 54]. While improving angular and spatial resolution allows for better characterization of fibre crossings, doing so drastically reduces the signal-to-noise ratio (SNR). The pronounced measurement noises tend to obscure and distort diagnostically relevant details of DW-MR signals [54], thereby necessitating application of the efficient and reliable tools of image denoising. One classic method to decrease MRI measurement noise is to average several (typically 3 to 10) measurements of the signal on the expectation that the random contributions will cancel out. In HARDI, however, an increased number of acquisitions would lead to prohibitive scan durations and alternative methods must be explored. A more efficient solution is to average similar samples in the same dataset. Thus, for example, since HARDI signals are effectively band-limited denoising can be accomplished by means of linear filtering. However, the spatial dependency of diffusivity in brain tissue makes it impossible to find a single set of linear filter parameters which is optimal for all types of diffusion signals. An adaptive filtering technique seems necessary. This thesis proposes a new type of non-local means (NLM) filtering which possesses the required adaptivity property. The proposed NLM filtering, as opposed

to similar methods in the field, includes in the average samples from all dimensions of the signal domain. Moreover, the weights are designed to be based solely on a sample's local orientational diffusion characteristics, are invariant to spatial rotations, and are independent of the particular sampling scheme in use. This thesis provides a detailed description of the proposed filtering procedure, as well as experimental results with synthetic and real-life data. It is demonstrated that the proposed filter has substantially better denoising capabilities as compared to a number of alternative methods.

# Chapter 2

# Literature Review

In virtually all realizations of MRI, attaining higher spatial resolution entails using longer acquisition times. Apart from being highly undesirable from the perspective of patients' comfort and compliance, longer acquisition times lead to motion-related artifacts, which are the main foe of diffusion MRI [52, Ch.2]. On the other hand, the reduction of acquisition time results in a loss of spatial resolution as well as in an amplification of measurement noises. Moreover, attaining high angular resolution in the recovery of the Ensemble Average Propagator (EAP) and EAP-related functions requires the use of strong encoding gradients, which has the undesirable effect of further decreasing the signal-to-noise ratio (SNR). Consequently, pronounced measurement noises tend to obscure and distort diagnostically relevant details of DW-MR signals [54], thereby necessitating application of the efficient and reliable tools of image denoising. In this chapter, we review a number of available filtering techniques developed specifically for HARDI denoising. First, however, prerequisites regarding the mathematical nature of the HARDI signal are provided.

## 2.1 HARDI signal description

In this section we define the HARDI signal in a purely mathematical manner, devoid of physical significance, and postpone the theory of signal acquisition until Chapter 3. It is of fundamental importance to consider the manifold to which the HARDI signal belongs, since proper filtering ought to take the signal domain into account. The HARDI signal $S(\mathbf{r}, \mathbf{u})$ can be made precise as a scalar function defined on the product space $\mathbb{R}^3 \times \mathbb{S}^2$, i.e. $S(\mathbf{r}, \mathbf{u}) : \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$. In practice, the HARDI signal is measured over a discrete domain. One illustration of sampling locations on the $\mathbb{R}^3 \times \mathbb{S}^2$ manifold is that of a discrete lattice of

discrete locations on the unit sphere (see Figure 3.6 in Chapter 3). Whereas this portrayal is particularly useful for visualizing denoising schemes based on the weighted averaging of samples, a number of alternative signal descriptions are important for visualizing other types of denoising approaches. One such description defines the HARDI signal as discrete in $\mathbb{R}^3$ and continuous in $\mathbb{S}^2$ - visualized as a discrete collection of continuous spherical functions $\{S_n(\mathbf{u})\}$ where $n$ is an indexing parameter and $S_n(\mathbf{u}) : \mathbb{S}^2 \rightarrow \mathbb{R}$. The first method of Section 2.4 took this view and proposed to denoise each spherical function independently. Alternatively, the HARDI signal can be defined as continuous in $\mathbb{R}^3$ and discrete in $\mathbb{S}^2$ - visualized as a discrete collection of continuous Euclidean images $\{S_k(\mathbf{r})\}$ where $k$ is an indexing parameter and $S_k(\mathbf{r}) : \mathbb{R}^3 \rightarrow \mathbb{R}$. The methods in Section 2.2 consider the images in this collection to be independent and denoise each one separately. However, the Euclidean images can also be coupled by combining them in a vector. This is done in both Section 2.5 and Section 2.4. It is then proposed to denoise the resulting vector-valued image $S(\mathbf{r}) : \mathbb{R}^3 \rightarrow \mathbb{R}^K$, where $K$ is the number of elements in the indexing set of $k$. Thus, whereas the methods in Section 2.2 smooth image intensities in the $\mathbb{R}^3$ domain, the methods in Sections 2.4 and 2.5 smooth sampled spherical functions in the $\mathbb{R}^3$ domain.

In addition to defining the HARDI signal domain, it is also important to consider the characteristics of HARDI measurement noise. The majority of filtering techniques available in the imaging sciences are based on an assumption of additive white Gaussian noise. Unfortunately, HARDI measurements follow a Rician distribution. Although it is true that the Rician distribution can be closely approximated by a Gaussian distribution for large SNR, HARDI often requires the use of strong diffusion-encoding gradients which cause profound signal attenuation, bringing signal amplitudes near the noise floor. In these circumstances, filtering techniques may benefit from taking into consideration the Rician nature of the measurements.

## 2.2 Individual DW-image denoising

The methods summarized in this section are designed to denoise classical MR images. In a straightforward approach, these methods can be applied individually to each DW-image composing the HARDI signal. However, this strategy ignores the manifold to which the signal belongs and, consequently, ignores pertinent information contained in the angular dependencies between samples. Nevertheless, the following paragraphs provide a compressed tour of available techniques relevant to MRI denoising.

The current arsenal of image denoising methods used in MRI is immense, which makes their complete classification a non-trivial task. For this reason, only three groups of denois-

ing methods which are germane to the present developments are mentioned below, while the reader is referred to the references for a more comprehensive literature review. The first group of denoising algorithms for MRI encompasses variational methods, which are implemented through the solution of partial differential equations (PDE) [12, 37, 41, 57]. Thus, for example, [41] suggests an adaptation of the classical anisotropic diffusion filter of [72] for noise reduction and enhancement of object boundaries in MRI. On the other hand, the denoising method of [57] is based on minimization of an original cost functional, whose associated gradient flow has the form of a fourth-order PDE. In [37], information from both the body coil image and surface coil image are incorporated in the form of data fidelity constraints. Finally, [12] introduces a *maximum-a-posteriori* (MAP) technique using a Rician noise model in combination with spatial regularization.

A second group of denoising methods takes advantage of the sparsifying properties of certain linear transforms [5, 8, 48, 49, 70, 74, 94, 101, 103, 107] (for a comprehensive review of such methods, the reader is also referred to [75]). Thus, for instance, the method of [70] is based on wavelet thresholding applied to squared-amplitude MR images, supplemented by an "unbiasing" of the scaling coefficients to account for the non-central chi-square distribution statistics. A different (robust) shrinkage scheme in the domain of a wavelet transform is proposed in [74]. Using a different line of arguments, the wavelet denoising method of [107] is applied to complex-valued MR images. Finally, in [5], the MR images are enhanced by means of a wavelet-domain bilateral filter.

A third group of image denoising algorithms is based on the concept of non-local means (NLM) filtering, which was originally proposed in [20–22], with later improvements reported in [18, 29]. As a general rule, NLM filters estimate a noise-free intensity of a given pixel as a weighted (linear) combination of the rest of the image pixels. Here, the weights of the linear combination are determined based on a *similarity measure* between the neighbourhoods of the "target" and "source" pixels. As a result, the performance of an NLM filter is largely defined by the optimality of a chosen similarity measure with respect to the properties of the image to be enhanced as well as those of the measurement noise. Thus, for example, under the conditions of additive white Gaussian (AWG) noise contamination, the above-referred NLM filters have been shown to outperform many variational and wavelet-based filters in terms of noise removal and the quality of edge preservation.

Motivated by the success of NLM filtering in general image processing, the works in [28, 99] have extended the Gaussian-mode NLM filters to MR imagery. Additional reports on the subject also include [34, 62, 63, 100], where the filters are applied to squared-magnitude MR images, followed by subtracting an estimation bias from the results thus obtained. Specific to DW-MRI, [34] has reported improved quality of anisotropy maps computed from the apparent diffusion coefficient and orientation distribution function.

## 2.3 Kernel smoothing on $\mathbb{S}^2$

Perhaps the the most famous denoising method applied in imaging sciences is Gaussian filtering, which reduces noise by convolving the noisy image with a Gaussian kernel. This method can also be reformulated as the solution to a diffusion equation. Namely, it can be shown that Gaussian kernel smoothing is equivalent to performing a diffusion process on the initial data where the time duration of diffusion is related to the kernel's effective support. Special care must be taken when extending diffusion smoothing to non-Euclidean space. Differential equations describing a diffusion process which is restricted to evolve on a particular manifold must use a definition of the derivative which takes the manifold's geometrical shape into consideration.

A number of methods have been proposed to extend diffusion smoothing to the $\mathbb{S}^2$ manifold. Whereas some approaches have attempted to numerically solve the differential equation for spherical diffusion, [24] and [26] proposed a solution using the equivalent operation of kernel smoothing on $\mathbb{S}^2$. Both of the above references analytically constructed the *Gauss-Weierstrass kernel* using spherical harmonics, albeit employing different derivations. It was shown that, analogous to the case in Euclidean space, spherical convolution with the Gauss-Weierstrass kernel is equivalent to performing a spherical diffusion process. The effective support of the kernel determines the degree of filtering and is related to the time duration of the diffusion process. In particular, large kernel support is used to filter low frequency signals and small support to filter high frequency signals. In application to HARDI denoising, Gauss-Weierstrass kernel smoothing can be independently performed on the spherical function at each spatial location. Unfortunately, the diverse types of spherical functions encountered in HARDI make it impossible to find a single, optimal kernel width.

Improved performance can be gained from kernel smoothing by regarding the technique as a weighted average, where weights are based on proximity. The average provides an approximation of the estimated sample's expected value. In the case of MRI statistics, the expected value estimate can then be used to approximate the signal's true amplitude by removing a bias as described in Section 4.4 (see also Section 3.3).

## 2.4 Variational denoising of diffusion MRI

Variational methods fit a continuous function to measurements via the minimization of an *energy functional* – an integral involving the unknown function. Noise is removed by regularization, a technique employed to overcome the ill-posedness of many reconstruction

problems. It is important to note that the numerical method used to minimize the functional must take into account the geometric shape of the manifold to which the solution belongs. In the case of the $\mathbb{R}^3 \times \mathbb{S}^2$ manifold, these numeric methods become complicated and, to simplify the problem, regularization is often performed in only one domain of the product space. The denoising approach proposed by McGraw et al. in [64] used two different variational frameworks to separately regularize the HARDI signal in the $\mathbb{S}^2$ domain and the $\mathbb{R}^3$ domain.

Regularization in the $\mathbb{S}^2$ domain was accomplished by *smoothing splines*, an adaptation of archetypical *interpolating splines*. Interpolating splines are piecewise-polynomial functions that minimize an energy functional with respect to additional interpolation constraints. The variational problem seeks to find a smooth function with low-valued derivatives and the specific derivatives included in the energy functional determine the type of spline minimizer. The method of smoothing splines, on the other hand, modifies this variational problem by relaxing the interpolation constraints and replacing them with a data infidelity penalty functional in the objective. Thus, by varying the weight of the penalty function, a trade-off between data fidelity and signal smoothness is introduced into the solution. It should be noted that one can invert the roles of the above functionals, viewing a data fidelity functional as the original objective and the spline functional as a roughness penalty.

More specifically, McGraw et al. used a weighted combination of the thin-plate spline and membrane spline. The energy functional of these splines measures the potential energy stored in a hybrid membrane due to a deformation from its relaxed state. In the absence of interpolation constraints or a data fidelity term, the minimizer is a constant function over the sphere. Introducing data fidelity into the objective is equivalent to attaching springs between the membrane and the observed data points. Therefore, the complete variational problem minimizes the potential energy in both the membrane and the springs in order to reach an equilibrium, while tuning parameters control the relative strength of the two opposing forces. Strong springs and a weak membrane produce a fluctuating minimizer which passes through the observed data, whereas weak springs and a firm membrane produce a smooth minimizing function which is allowed to deviate significantly from the observed data. In practice, one tries to find parameters which result in a smooth solution that is close to the observed data but not fitted to the noise. For example, stronger springs and a weaker membrane are used to fit data whose noise-free signal is known to contain high frequency components. On the other hand, weaker springs and a stronger membrane are used to fit the noisy observations of a low frequency signal. However, the diverse types of spherical functions encountered in diffusion imaging makes it impossible to find a single set of optimal tuning parameters with respect to the underlying function regularity. Apart

from the underlying signal, the variance of measurement noise also informs the choice of tuning parameters. In the case that measurement noise has small variance, strong springs are used, while weak springs are used with large noise variance. Unfortunately, the variance of MRI measurement noise fluctuates with signal amplitude (see Figure 3.3) making it impossible to find a single set of optimal tuning parameters with respect to noise variance.

Among the various numerical techniques available to solve spline variational problems, McGraw et al. used a standard method of finite element methods (FEM). In particular, the global coordinates of the FEM system were chosen to be spherical and the domain was discretized into triangular elements using icosahedron sampling. Additionally, the fifth order element shape functions given by Dhatt and Touzot [35] were used, as this element guarantees $C1$ continuity across triangles.

Smoothing in the spatial domain $\mathbb{R}^3$ was accomplished via total variational (TV) regularization. TV was introduced by Rudin et al [79] and has proven to be successful in removing noise while preserving edges. The variational minimization problem incorporates a data fidelity term which brings the solution close to the observed data in the $L_2$ sense and a TV roughness penalty term which encourages the solution's gradient to be small. Blomgren and Chan have investigated extending TV regularization to vector-valued images [17]. Considering the HARDI signal as continuous in $\mathbb{R}^3$ and discrete in $\mathbb{S}^2$, a vectorfield with vectors corresponding to sampled spherical functions, McGraw et al. followed Blomgren and Chan's approach with two modifications. Blomgren and Chan proposed the TV norm to be defined as the $l^2$ norm of a vector containing each scalar image's classical TV norm. McGraw et al. modified the definition to use an $l^1$ norm instead and achieved a decoupling which would permit a form of TV denoising to be performed on each DW-image independently. However, McGraw et al. also introduced an adaptive weighting on the TV regularization term which re-coupled the denoising problem. The TV norm was weighted adaptively according to the generalized anisotropy index defined in [71], permitting strong regularization in regions where sampled spherical functions have homogeneous anisotropy and stopping denoising on anisotropic edges. The generalized anisotropic index is computationally more efficient to implement than the coupling term found in Blomgren and Chan's approach. The final variational problem was solved using a variational gradient descent approach with the resulting partial differential equation being solved using "fixed-point lagged diffusivity.".

Both of the aforementioned variational approaches from [64] make use of a regularization term. Regularization compensates for missing degrees of freedom in reconstruction problems by cleverly including an a priori assumption about the solution. Formally, regularization can be based on a Bayesian estimation framework, which requires pre-defining an a priori statistical model for the unknown image of interest. Since different prior models

result in different estimation outcomes, the choice of prior is crucial. However, in the field of medical imaging, there is no consensus as to which type of image prior should be used and there does not seem to be any objective means to assess the validity of one prior over another. Moreover, the incorrect definition of a prior model can result in a distorted appearance of diagnostically relevant features, which could potentially lead to misdiagnosis. This is why, when little is known about the nature of the image of interest (as is typically the case in medical imaging due to inter-subject variability and the intrinsic intricacy of biological tissues), it makes sense to design image reconstruction procedures to be as independent as possible from a priori assumptions.

## 2.5 Total variation denoising of the ADC

The variational approach proposed by Kim et al. in reference [56] is similar to the spatial regularization method reviewed in the second half of the previous section. However, instead of denoising the raw HARDI signal, Kim et al. proposed to denoise the Apparent Diffusion Coefficient (ADC) by fitting an ADC model (see Equation 3.34) to the observed HARDI data. The adopted model was chosen to be continuous in the spatial domain $\mathbb{R}^3$, yet discrete in the spherical domain $\mathbb{S}^2$. In this case, it is helpful to view both the ADC parameter and corresponding image formation model as a vectorfield over $\mathbb{R}^3$; at each spatial location exists a vector of samples taken from an underlying spherical function. Kim et al. proposed a vector total variation approach for ADC regularization in the $\mathbb{R}^3$ domain. The recovered ADC can subsequently be substituted into the image formation model to obtain the denoised HARDI data.

The energy functional to be minimized was composed of three terms. The first was a data fidelity term forcing the ADC model to be close to the observed data. Data fidelity was constructed with an $L_1$ norm minimization rather than the more standard $L_2$ norm minimization. The second term was a vectorial TV regularization term, first introduced by Blomgren and Chan for color images [17]. Kim et al. proposed to denoise the ADC and, accordingly, regularization was implemented as the vectorial TV norm of the ADC parameter. The final term in the energy functional was a constraint term which ensured that the minimizing ADC solution only admit positive values. Negative ADC values would imply signal amplification due to diffusion, a physical impossibility. The constraint term was realized by an adaptation of the log barrier method. The minimizer was found numerically using the gradient descent method, with the resulting partial differential equations being solved by finite difference methods. Referring the reader to the discussion in the last paragraph of the previous section, it is briefly noted here that although regularization rec-

tifies the ill-posedness of the reconstruction problem, there is some ambiguity as to which regularization term is most correct and the degree to which an incorrect choice will distort the appearance of diagnostically relevant features.

## 2.6 NLM for diffusion MRI

The non-local means filter (NLM), which was originally proposed in [20–22], estimates the noise-free intensity of a given "source" sample as a weighted average of other "target" samples in the image. Whereas kernel smoothing schemes restrict large weights to be assigned to target samples in the vicinity of the source sample, non-local means makes no such restriction. Instead, the target sample's weight in the average is decided solely on a similarity measure between source and target neighbourhoods.

Nicolas Wiest-Daessle et al. proposed a number of NLM variants for the denoising of diffusion-weighted MRI [99]. The filter most relevant to HARDI treats the signal as a discretized vectorfield - i.e. at each discrete location in space is a vector of samples taken from an underlying spherical function. Therefore, instead of estimating a source *sample* as a weighted average of target *samples*, the proposed filter estimates a source *vector* as a weighted average of target *vectors*. Correspondingly, each target vector's contribution to the average is determined by a similarity measure between vector neighbourhoods. Thus, Wiest-Daessle et al. compare neighbourhoods which are localized in $\mathbb{R}^3$ but include all information from $\mathbb{S}^2$. The method, however, seems slightly restrictive when considered from the viewpoint of estimating a single sample. For example, when estimating a source sample, only target samples with the same spatial orientation are included in the average. In other words, it is inherently assumed that target samples with a different spatial orientation are not similar to the source sample. The above considerations indicate that it may be beneficial to investigate comparing neighbourhoods localized in $\mathbb{S}^2$ in order to increase the number of relevant target samples included in the source sample estimation.

# Chapter 3

# Theory

The purpose of this chapter is to provide the theoretical background of diffusion-weighted (DW) magnetic resonance imaging (MRI). For completeness, a brief overview of classical MRI theory is given in Section 3.1. Continuing, the developments of DW-MRI are provided in Section 3.2. In particular, high angular resolution diffusion imaging (HARDI) acquisition is described in Section 3.2.4, which provides the theory necessary to understand in what sense neighbouring HARDI samples are "close" to each other. HARDI measurement noise is described in Section 3.3. Finally, spherical harmonics, an important tool for HARDI signal analysis, are introduced in Section 3.4.

## 3.1 Physics of magnetic resonance imaging

Living tissue in the human body is predominantly composed of water. In fact, the water content of the human body is roughly 60%. MRI is able to leverage the predominance of water in tissue by measuring its behaviour as a function of spatial location. By subjecting water molecules to a sequence of magnetic pulses and measuring the response, data is acquired permitting the detailed display of a subject's internal structure. This section contains a brief and simplified introduction to the physics of the MR imaging modality which will be beneficial in understanding the future sections concerning diffusion-weighted imaging. The information provided in this section is largely a summary of the references [25] and [46].

### 3.1.1 The spin

The water molecule, which is essential to the functioning of MRI, is composed of three atoms: one oxygen atom and two hydrogen atoms. The MRI scanner manipulates hydrogen atoms, which makes hydrogen's atomic properties a fundamental subject of MRI. According to the theory of quantum mechanics, each hydrogen nuclei posses the fundamental property of spin. Spin was classically conceived as the rotation of an electrically charged particle about an axis which induces a magnetic moment. Moreover, the direct relationship between the particle's angular momentum $\mathbf{J}$ (or spin property) and the induced magnetic moment $\boldsymbol{\mu}$ is given by

$$\boldsymbol{\mu} = \gamma \mathbf{J}, \tag{3.1}$$

where $\gamma$ is the gyromagnetic constant, a proportionality constant with units rad/s/T, found experimentally.

Although it is true that the atomic property of spin obeys the same mathematical laws as angular momentum, a number of additional properties have been discovered which make spin categorically different from orbital angular momentum. However, for our purposes, it will be convenient to continue considering spin as the classical idea of a charged particle rotating about an axis. It is important to note that due to the close relation between spin and magnetic moment, it is standard to refer to the hydrogen magnetic dipoles $\boldsymbol{\mu}$ as *'spins'*. The following sections explain how spins are harmonized in order to generate a net magnetic effect which is measurable.

### 3.1.2 Larmor precession and alignment

Under normal circumstances, a collection of spins in a sample have no fixed orientation; the orientations are randomly distributed throughout the sample and the magnetic effect of any particular spin is counteracted by the magnetic effects of other spins in the sample. However, it is possible to use an external magnetic field to align the spins in the same direction. In the presence of an external magnetic field, $\mathbf{B}$, a torque is exerted on the spins akin to the torque experienced by a current loop when exposed to an external magnetic field. The torque $\mathbf{N}$ on a spin $\boldsymbol{\mu}$ is given by

$$\mathbf{N} = \boldsymbol{\mu} \times \mathbf{B}. \tag{3.2}$$

This torque requires that the hydrogen's total angular momentum $\mathbf{J}$ change according to

$$\frac{d\mathbf{J}}{dt} = \mathbf{N} \tag{3.3}$$

Upon substituting (3.1) and (3.2). into (3.3), it is observed that:

$$\frac{d\boldsymbol{\mu}}{dt} = \gamma\boldsymbol{\mu} \times \mathbf{B}. \tag{3.4}$$

If the external magnetic field $\mathbf{B}$ is along the z-axis and is constant in time, denoted $\mathbf{B_0} = |\mathbf{B_0}|\hat{z}$, then the solution to (3.4) is the well-known result of *precession*. Thus, the external magnetic field causes the spin to precess about $\mathbf{B_0}$ (see Figure 3.1) analogous to the precession of a spinning top about a gravitational field. The precession solution can be written in Cartesian form as

$$\begin{aligned}
\mu_x(t) &= \mu_x(0)\cos(\omega_0 t) + \mu_y(0)\sin(\omega_0 t) \\
\mu_y(t) &= \mu_y(0)\cos(\omega_0 t) - \mu_x(0)\sin(\omega_0 t) \\
\mu_z(t) &= \mu_z(0)
\end{aligned} \tag{3.5}$$

It is important to note that the angular frequency $\omega_0$ with which the spin precesses is given by the *Larmor equation*

$$\omega_0 = \gamma|\mathbf{B_0}|, \tag{3.6}$$

and is referred to as the *Larmor frequency*. An interesting consequence of (3.6) is that if $\mathbf{B_0}$ is not spatially uniform, then spins located at positions with greater field intensity will precess with increased frequency.

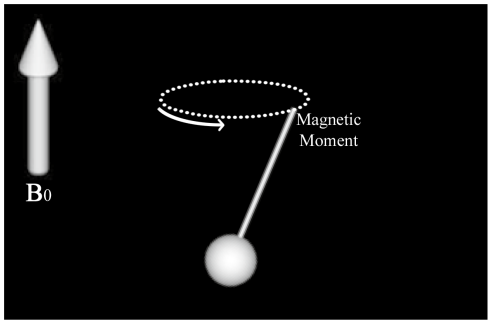



Figure 3.1: A spin precessing about $\mathbf{B_0}$

Although not evident from (3.5), in reality, a spin cannot precess about $\mathbf{B_0}$ indefinitely. While precessing, the spin dissipates energy into the surrounding environment by means of vibration and increased temperature. This process causes all spins to adjust exponentially into a state of thermal equilibrium and alignment with $\mathbf{B_0}$ (see Figure 3.2). This phenomenon, known as relaxation, is discussed in Section 3.1.5.

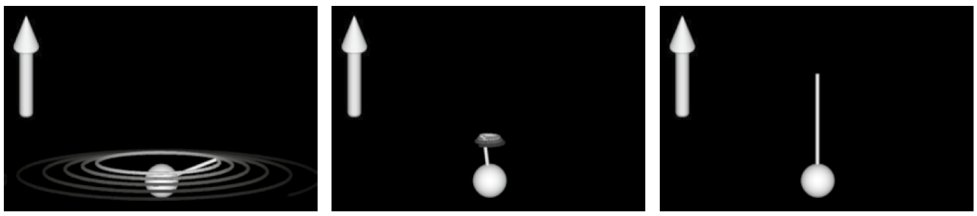


Figure 3.2: The process of a spin aligning with $\mathbf{B_0}$ [52]

### 3.1.3 Resonance and in-phase precession

While the application of a static magnetic field $\mathbf{B_0}$ induces spin precession, other types of magnetic fields induce different spin behaviour. In particular, an oscillating magnetic field applied perpendicular to $\mathbf{B_0}$ rotates a spin out of alignment with $\mathbf{B_0}$. This phenomenon is referred to as *resonance* and is noticeable in the name of magnetic resonance imaging.

The oscillating field, or *rf pulse*, is, as the name suggests, only applied for a short duration and in the radio frequency range. Analyzing the effects of an rf pulse is best accomplished using a rotating reference frame and more details can be found in references [25, 46]. The analysis shows that rf pulses are characterized by their *flip angle*, the degree to which they tip a spin. For example, a 90° pulse tips the spin perpendicular to $\mathbf{B_0}$ whereas a 180° pulse turns the spin upside-down. The flip angle is increased by increasing the amplitude or duration of the rf pulse.

One important application of the rf pulse is the process of *excitation*. A spin which has reached a state of alignment with $\mathbf{B_0}$ can be tipped into the transverse plane via application of a 90° pulse. Once out of alignment, the spin again begins to precess about $\mathbf{B_0}$ (see Figure 3.3). All spins which are simultaneously excited precess about $\mathbf{B_0}$ simultaneously and in phase.

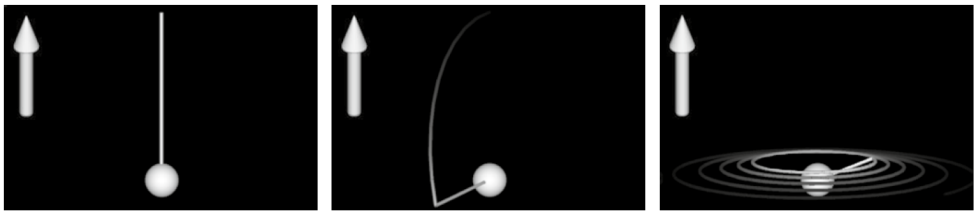


Figure 3.3: The process of excitation [52]

It is important to emphasize that only spins precessing with the same frequency as the

rf pulse experience the excitation torque. If a spin's precession frequency differs from the rf oscillation frequency, not much happens except a little wobbling around $\mathbf{B_0}$. In the case of a spatially non-uniform $\mathbf{B_0}$, the spin precession frequency varies with spatial location (see Equation 3.6) and it is possible to selectively excite specific spins.

### 3.1.4 Idealized sample spin density measurement

An excited spin is essentially a rotating magnetic dipole which, according to Faraday's law of induction, can be detected through the voltage induced in a receiver coil (see Figure 3.4). Faraday's law states that a fluctuating magnetic flux through a closed circuit induces an electromagnetic force (voltage) in that circuit. The voltage induced is proportional to the rate of change of magnetic flux. Formally, Faraday's law is stated

$$V(t) = -\frac{d\Phi(t)}{dt}, \tag{3.7}$$

where $V(t)$ is the voltage and $\Phi(t)$ is the magnetic flux.

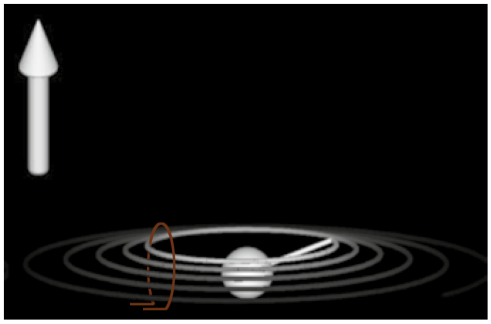


Figure 3.4: Magnetic dipole cutting across a receiver coil

Due to the initially random orientations of the dipoles, spins will precess out of phase on first application of an external magnetization field and no noticeable signal will be detectable. Spins are brought to precess in phase through alignment and excitation, after which spin dipoles constructively interfere to induce a measurable voltage signal. Naturally, the frequency of the signal is equal to the Larmor frequency.

A more precise development can be made by considering the collective magnetization of all the spins in a volume $V$, which is known as the *magnetization vector* $\mathbf{M}$

$$\mathbf{M} = \frac{1}{V} \sum_{\substack{protons \\ i \in \ in \ V}} \boldsymbol{\mu}_i. \tag{3.8}$$

It is standard to make a simplifying assumption that the applied external magnetization field is in the $\hat{z}$ direction (i.e.. $\mathbf{B_0} = |\mathbf{B_0}|\hat{z}$). In this case, only the transverse components of the magnetization vector cut across the receiver coil to contribute to the induced voltage. The transverse component of $\mathbf{M}$ is written

$$\mathbf{M}_\perp = M_x\,\hat{x} + M_y\,\hat{y}. \tag{3.9}$$

After a number of idealizations and approximations have been made, the total voltage induced in the receiver coil can be derived to be proportional to [46]

$$s(t) \propto \int |\mathbf{M}_\perp(\mathbf{r}, t)| \sin(\omega_0 t) d\mathbf{r}, \tag{3.10}$$

where $\omega_0$ is the Larmor frequency. The high frequency oscillations are removed by a process of demodulation and the resulting signal depends only on the transverse magnetization vector, which is proportional to spin density. Thus, by appropriately normalizing the demodulated signal, the spin density of the sample can be determined. It should be noted that, rather than measure the spin density of an entire sample, it is also possible to measure the spin density of a specific voxel. However, this requires the use of gradient magnetic fields, the subject of Section 3.1.6.

### 3.1.5 Relaxation and the Bloch equation

The solution expressed in (3.5) seems to indicate that spins, and their corresponding magnetization vector, should oscillate indefinitely in the presence of a static external magnetic field. The analysis, however, ignores interactions between the spins and their surrounding environment. Rather than precess indefinitely, spin interactions cause the magnetization vector to return to its thermal equilibrium position. Although properly understood in the context of quantum mechanics, some understanding of these interactions can be gained from a classical perspective. The phenomena are captured in the *T1, T2, and T2* relaxation* effects.

T1 relaxation describes the recovery of the magnetization vector's $M_z$ component to thermal equilibrium as spins dissipate extra energy, acquired from excitation, into the surrounding environment. T2 relaxation, on the other hand, describes the decay of $\mathbf{M}_\perp$ due to an irreversible dephasing of the spins. Specifically, T2 relaxation is a loss of synchronization which occurs as each spin's Larmor frequency changes on account of neighbouring spins' oscillating magnetic fields. It should be noted that the T1 and T2 relaxation are separate processes transpiring over different timespans (i.e.. relaxation is not simply a rotation of

the magnetization vector from the transverse plane to the $\hat{z}$ direction) and it is typical for $\mathbf{M}_\perp$ to decay faster than the recovery of $M_z$. T1 and T2 relaxation times are determined by the substance's biochemical composition, making both important for internal structure analysis. By altering pulse sequence timing parameters, it is possible to allow either T1 or T2 relaxation to dominate the received signal producing the T1-weighted and T2-weighted images, respectively. Alternatively, proton density (PD) contrast is achieved by limiting both relaxation effects.

The effect of T1 and T2 relaxation on the Magnetization vector are captured in the phenomenological *Bloch equation* [25, 46]

$$\frac{d\mathbf{M}}{dt} = \gamma(\mathbf{M} \times \mathbf{B_0}) + \frac{1}{T_1}(M_0 - M_z)\hat{z} - \frac{1}{T_2}\mathbf{M}_\perp. \tag{3.11}$$

In the case that $\mathbf{B_0} = |\mathbf{B_0}|\hat{z}$, the solution to the Bloch Equation is given by

$$\begin{aligned} M_x(t) &= e^{-t/T_2}[M_x(0)\cos(\omega_0 t) + M_y(0)\sin(\omega_0 t)] \\ M_y(t) &= e^{-t/T_2}[M_y(0)\cos(\omega_0 t) - M_x(0)\sin(\omega_0 t)] \\ M_z(t) &= M_z(0)e^{-t/T_1} + M_0(1 - e^{-t/T_1}) \end{aligned} \tag{3.12}$$

Note the similarity to the pure precession solution of (3.5) with the addition of exponential decays.

An additional relaxation effect, known as T2* relaxation, also occurs due to spin dephasing. Small inhomogeneities in $\mathbf{B_0}$ cause small variances between spin Larmor frequencies which results in dephasing. However, unlike T2 dephasing, T2* dephasing is reversible. Hahn's spin echo method is capable of recovering spin coherence via a *refocusing pulse*. The refocusing pulse is an rf pulse with 180° flip angle, which inverses the spins' orientation in the transverse plane. Applied at a time $\tau$ after excitation, the refocusing pulse puts the faster precessing spins behind the slower precessing spins. At a time $2\tau$ after excitation the fast spins catch up to the slow spins and all spins are back in phase. The following common analogy provides a helpful illustration of the spin echo method. Imagine a race, where every runner starts at the line (phase coherence after excitation) and runs with different but constant speeds (different Larmor frequencies). After a time $\tau$ all runners are at different positions (dephasing). However, if they simultaneously reverse directions (refocusing pulse), maintaining their respective speeds, then at time $2\tau$ all runners will again be in line.

### 3.1.6 Gradient magnetic fields and spatial resolution

The application of nuclear magnetic resonance to internal structure analysis requires spatial resolution - the ability to recover the magnetization vector at every point in space. Discrimination of spin characteristics with respect to spatial position is achieved by varying the Larmor frequency as a function of location. According to the Larmor Equation 3.6, the addition of a *persistent* gradient magnetic field will gradually increase the precessional *frequency* of the spins along the gradient direction. On the other hand, a *pulsed* gradient magnetic field changes the *phase* of the spins along the gradient direction. The manipulations of spin frequency and phase allow for slice selection and phase encoding, respectively.

#### Slice selection

Slice selection is a procedure used to selectively excite a single 2D plane of spins in the sample. In accordance with the Larmor Equation 3.6, applying a linear gradient magnetic field causes a linear spatial change in spin angular frequency along the gradient direction. Thus, applying an rf pulse of a specific frequency will excite only those spins in spatial locations with matching Larmor frequencies. For example, if the gradient magnetic field direction is $\hat{z}$, then at each z location is a transverse plane with spins with the same Larmor frequency (it is helpful to think of the entire volume as a stack of transverse planes, each having a different Larmor frequency). By setting the rf pulse to match only one plane's Larmor frequency, only that slice of spins are excited. Spins not in the excited plane remain aligned with $\mathbf{B_0}$ and do not contribute to the induced signal in the receiver coil.

#### Phase encoding and K space

Following slice selection, phase encoding enables the measurement of the magnetization vector at each point in the selected slice, which we again consider to be the x-y transverse plane. In order to precisely articulate the theory of phase encoding, it is first necessary to state and discuss the mathematical equations resulting from the demodulation of the received signal (3.10). Following demodulation, the signal received from real and quadrature channels can be written in complex representation as (see [46] for more details)

$$s(t) \propto \int |\mathbf{M}_{\perp}(\mathbf{r}, t)| e^{i(\Omega t - \phi(\mathbf{r},t))} d\mathbf{r}. \tag{3.13}$$

The symbol $\Omega$ denotes the demodulation frequency and the angle $\phi(\mathbf{r}, t)$ is the accumulated phase of the spins at location $\mathbf{r}$. The accumulated phase is calculated by integrating the

spin frequency with respect to time

$$\phi(\mathbf{r}, t) = \int_0^t \omega(\mathbf{r}, \tau) d\tau, \tag{3.14}$$

where frequency $\omega(\mathbf{r}, t)$ is determined by Larmor Equation 3.6. In the absence of a gradient, a spatially uniform external magnetic field produces a spatially uniform accumulated phase of $\phi(\mathbf{r}, t) = \omega_0 t$ where $\omega_0$ is the Larmor frequency. However, if the external field is non-uniform, then the accumulated phase depends on location $\mathbf{r}$

$$\phi(\mathbf{r}, t) = \int_0^t \omega_0 + \omega'(\mathbf{r}, \tau) d\tau = \omega_0 t + \phi'(\mathbf{r}, t). \tag{3.15}$$

The additional terms $\omega'(\mathbf{r}, t)$ and $\phi'(\mathbf{r}, t)$ bring to attention the change in frequency and therefore accumulated phase at spatial locations where the magnetic field differs from $\mathbf{B_0}$.

Phase encoding involves the application of pulsed gradients $G_x$ and $G_y$ to vary the intensity of the magnetic field along the $\hat{x}$ and $\hat{y}$ directions of the transverse plane. The deviation from $\mathbf{B_0}$ results in the spins at each (x,y) position accumulating a different phase dependent on position

$$\phi(x, y, t) = \omega_0 t + 2\pi(k_x x + k_y y) \tag{3.16}$$

$$k_x = \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau \tag{3.17}$$

$$k_y = \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau \tag{3.18}$$

With a demodulation frequency $\Omega = \omega_0$ and accumulated phase given by (3.16), the received signal (3.13) can be written

$$s(t) \propto \int |\mathbf{M}_\perp(x, y)| e^{-i2\pi(k_x x + k_y y)} dx dy, \tag{3.19}$$

commonly rewritten

$$F(k_x, k_y) = \int f(x, y) e^{-i2\pi(k_x x + k_y y)} dxdy, \tag{3.20}$$

where $F(k_x, k_y)$ denotes the received signal and f(x,y) represents $|\mathbf{M}_\perp(x, y)|$.

Inspection of (3.20) shows that the mathematical recovery of $f(x, y)$ is possible using the inverse Fourier transform. In practice, K space is sampled at a sufficient number of points to recover $f(x, y)$ (discretized) using the inverse fast Fourier transform. The pixel intensity at each location in the recovered image $f(x, y)$ is dictated by the spin density, T1 relaxation, and T2 relaxation specific to that location.

Ideally, $f(x, y)$ should be recovered as a real-valued image, however, practical limitations result in $f(x, y)$ being complex-valued. This can occur, for example, if there is any error in matching $\Omega$ to $\omega_0$. In this case, the signal measured in k-space is

$$F(k_x, k_y) = \int f(x, y) e^{-i2\pi(k_x x + k_y y) + i\theta} dxdy = \int f(x, y) e^{i\theta} e^{-i2\pi(k_x x + k_y y)} dxdy, \tag{3.21}$$

where $\theta$ is a global phase shift equal to $(\Omega - \omega_0)$. Taking the inverse Fourier transform recovers the complex quantity $f(x, y)e^{i\theta}$. This problem is overcome by computing the MR magnitude image

$$|f(x, y)| = \sqrt{\Re\{f(x, y)e^{i\theta}\}^2 + \Im\{f(x, y)e^{i\theta}\}^2} \tag{3.22}$$

or simply,

$$M = \sqrt{(A\cos(\theta))^2 + (A\sin(\theta))^2}. \tag{3.23}$$

The above equation is instrumental in understanding the statistical nature of measurement noises in MR magnitude images. An example of an MR magnitude image is shown in Figure 3.5.

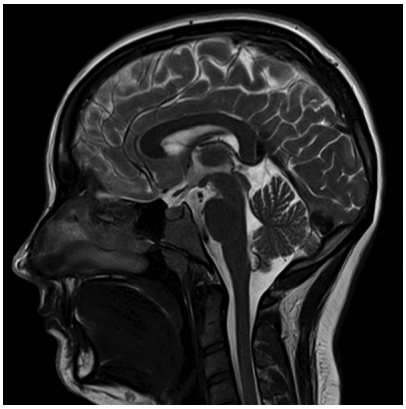


Figure 3.5: MR image of the human brain

## 3.2 Diffusion-weighted magnetic resonance imaging

Despite the importance of traditional T1,T2, and PD contrasts to clinical diagnosis, they have been unable to demonstrate effectiveness in delineating the morphological structure of the white matter – the information which is known to be related to a wide spectrum of brain-related disorders [7, 36, 38, 69, 78, 84, 105, 106]. It is only with the recent advent of *diffusion-weighted MRI* (DW-MRI) that scientists have been able to perform quantitative measurements of the diffusivity of white matter, making possible the structural delineation of neural fibre tracts in the human brain. This section outlines developments in the DW-MRI field which lead to the establishment of high angular resolution diffusion imaging (HARDI).

### 3.2.1 Diffusion weighted imaging

As the name suggests, diffusion weighted imaging (DWI) is an MR imaging modality in which the image voxel intensity is weighted by the strength of diffusion occurring in its vicinity. The weighting is accomplished using the spin echo method of Hahn, discussed in Section 3.1.5. Although Hahn recognized that spin echo was sensitive to diffusion, it was Carr and Purcell who developed the initial framework for making quantitative measurements of the diffusion coefficient. Later, Stejskal and Tanner introduced a novel Pulsed Gradient Spin Echo (PGSE) experiment which greatly improved upon previous techniques and remains the basis of the protocols implemented by modern scanners [83].

DWI employs two *diffusion-encoding gradients*, both denoted by $\mathbf{g}$. The first gradient

is applied prior to the refocusing pulse and the second gradient afterwards. The first encoding gradient intentionally causes T2* dephasing along the gradient's direction. The spin echo method is then used in conjunction with a second encoding gradient to reverse the dephasing. However, diffusion occurring in the direction of the encoding gradients during the time between their application, known as the *mixing time* $\Delta$, may impair the process of phase correction. Returning to the race analogy of Section 3.1.5, imagine a runner who, at the beginning of the race, initially jogs slowly (during the first encoding gradient, spin is in a position with low gradient field intensity), yet after reversing directions begins to sprint quickly (after refocusing and during the second encoding gradient, the spin has diffused into a high gradient field intensity). Similarly, another runner might initially sprint quickly and after reversing directions jog slowly. It is evident that the second runner will not be able to catch up with the first runner. This analogy illustrates how diffusion works against the spin echo method, inhibiting the refocusing of spin dephasing and causing an attenuation of the received signal. Yet, it is the measurement of signal attenuation which facilitates the determination of the diffusion coefficient.

Many of the early mathematical frameworks relating signal attenuation to the diffusion coefficient were specific to the particular diffusion-encoding protocol in use. However, an alternative and more general mathematical explanation was given by Torrey in 1956. Using the Einstein relation, Torrey generalized the phenomenological Bloch Equation (3.11) to include attenuation effects due to isotropic *free Gaussian diffusion* [87]

$$\frac{d\mathbf{M}}{dt} = \gamma(\mathbf{M} \times \mathbf{B_0}) + \frac{1}{T_1}(M_0 - M_z)\hat{z} - \frac{1}{T_2}\mathbf{M}_\perp + D\nabla^2\mathbf{M}, \tag{3.24}$$

where $D$ is the diffusion coefficient from Fick's Law. More generally, $D$ is referred to as the *apparent diffusion coefficient* (ADC), a name indicating that many materials do not permit the free Gaussian diffusion assumed in Torrey's derivation. Many tissues have a microstructure which inhibits spins along their random walks. These restrictions result in the spins' mean squared displacement per unit time being lower than that observed in free water. Consequently, the measured diffusion coefficient in these materials is also lower than what would be observed if the material truly followed free Gaussian diffusion.

The solution to the above Bloch-Torrey Equation when $\mathbf{B_0} = |\mathbf{B_0}|\hat{z}$ admits a transverse magnetization vector with an additional term accounting for attenuation due to diffusion

$$|\mathbf{M}_\perp^D(\mathbf{r}, t)| = |\mathbf{M}_\perp(\mathbf{r}, t)|e^{-bD(\mathbf{r})}. \tag{3.25}$$

This is the value obtained from K space reconstruction and a single voxel in the reconstructed image is commonly rewritten in the notation

$$S_D = S_0\, e^{-bD}, \tag{3.26}$$

with $S_0$ absorbing effects due to spin density, T1 relaxation, and T2 relaxation. The *b-value* appearing in the exponential depends on parameters specific to the diffusion-encoding protocol in use. For instance, Stejskal and Tanner's PGSE experiment admits a b-value of $b = (\gamma\delta|\mathbf{g}|)^2(\Delta - \frac{\delta}{3})$, where $\delta$ is the duration of the diffusion-encoding pulse [83]. As is evident from (3.26), the choice of b-value has an effect on the signal-to-noise ratio (SNR). Large b-values increase signal attenuation and lead to the dominance of measurement noise. Apart from b-value, the ADC at each spatial location also has an effect on the attenuation of each pixel in the DW-image. Recovering the ADC is the main goal of DWI, since it is the value which quantifies the strength of diffusion occurring local to the spatial location of a pixel. Toward this end, the value of $S_0$ is determined from an image acquired with no diffusion-encoding gradient ($b = 0$) and eliminated from $S_D$ by dividing $S_D/S_0$. The ADC is then recovered using the relation

$$D = -\frac{log(S_0/S_D)}{b}. \tag{3.27}$$

### 3.2.2 Diffusion tensor imaging

Diffusion weighted imaging is based on an assumption of *isotropic* diffusion. However, it was later found that the ADC is direction dependent. Namely, the microstructure of biological tissue often restricts diffusion in a direction dependent manner. For instance, muscle tissue and cerebral white matter limit diffusion to be along fibre bundles [14, 27, 47, 82]. In response, diffusion tensor imaging (DTI) [2, 9, 14, 15, 53, 66, 73, 98], has been developed for the measurement of *anisotropic* Gaussian diffusion and has found a broad range of applications in fibre tractography [11, 13, 59–61, 81].

In a 1965 paper, Stejskal investigated the effects of anisotropic diffusion on the ADC in a PGSE framework [82]. Using a tensor description of diffusion, Stejskal modified the Bloch equation to include the effects of anisotropic diffusion

$$\frac{d\mathbf{M}}{dt} = \gamma(\mathbf{M} \times \mathbf{B_0}) + \frac{1}{T_1}(M_0 - M_z)\hat{z} - \frac{1}{T_2}\mathbf{M}_\perp + \nabla^T \cdot \mathbf{D} \cdot \nabla\mathbf{M}, \tag{3.28}$$

where the diffusion tensor $\mathbf{D}$ is a matrix parametrizing Fick's Law for anisotropic Gaussian diffusion. The solution to the above equation when $\mathbf{B_0} = |\mathbf{B_0}|\hat{z}$ admits a transverse magnetization vector which is dependant on both spatial location $\mathbf{r}$ and the spatial orientation of the diffusion-encoding gradient $\mathbf{u} = \mathbf{g}/|\mathbf{g}|$

$$|\mathbf{M}_\perp^\mathbf{D}(\mathbf{r}, \mathbf{u}, t)| = |\mathbf{M}_\perp(\mathbf{r}, t)|e^{-b\mathbf{u}^T\mathbf{D}(\mathbf{r})\mathbf{u}}. \tag{3.29}$$

Consequently, the value of a specific voxel in an image reconstructed from K space not only depends on spatial location, but also on the orientation of the diffusion-encoding gradient. It is common to consider the possible values of a single voxel in the reconstructed image as a spherical function written in the notation

$$S_{\mathbf{D}}(\mathbf{u}) = S_0\, e^{-b\mathbf{u}^T\mathbf{D}\mathbf{u}}. \tag{3.30}$$

Note that the value of the ADC (in this case equal to $\mathbf{u}^T\mathbf{D}\mathbf{u}$) is also dependent on spatial orientation $\mathbf{u}$ and, due to the diffusion tensor description, is ellipsoidal.

Although anisotropic diffusion was addressed by Stejskal in 1965, it was not until 1994 that a method was provided by Basser to estimate the diffusion tensor using linear regression [9]. The estimation requires the measurement of at least seven independent MR images - one in the absence of diffusion encoding to capture $S_0$ and six more DW-images to fully specify $\mathbf{D}$. Typically, the six gradients are restricted to the same b-value and only the spatial orientation is changed.

Using the diffusion tensor, one can describe the spin diffusion along three orthogonal directions, with the principal eigenvector indicating the orientation of the underlying fibre tract. However, the implicit assumption of anisotropic Gaussian diffusion limits the ADC to a single orientational maximum. Consequently, proper functioning of DTI requires each voxel to contain no more than a single fibre bundle and severe underestimation artifacts occur in spatial locations where fibres cross, touch upon each other, or diverge [3, 4, 39, 40, 92, 93, 98].

### 3.2.3 Diffusion spectral imaging

Biological tissue often embodies complicated diffusion processes which differ significantly from the isotropic Gaussian model assumed by DWI and the anisotropic Gaussian model assumed by DTI. Diffusion spectral imaging (DSI) provides a model independent approach to characterizing complex types of diffusion. As with DTI, the foundation for DSI was laid by Stejskal. In the same paper where he provided the Bloch equations modified for anisotropic diffusion, Stejskal also included the development of a more general description of the diffusion process - the *Ensemble Average Propagator* (EAP). The EAP is a probability density function quantifying the probability of a spin to be found in a volume $\mathbf{r}\,d\mathbf{r}$ after a given diffusion time. Since this probability depends on the local microstructure of interrogated tissue, the EAP is essential for understanding tissue composition and organization.

Central to DSI is the *wave-vector*, which is related to a number of the encoding gradient parameters and defined to be equal to

$$\mathbf{q} = \frac{\gamma\,\delta\,\mathbf{g}}{2\pi}, \tag{3.31}$$

where $\gamma$ is the gyromagnetic constant, $\delta$ is the duration of the diffusion-encoding gradients, and $\mathbf{g}$ is a vector representing the magnitude and orientation of the encoding gradients. The value $S(\mathbf{q})$ of each voxel in an image recovered from K space can be written as a function of $\mathbf{q}$, which is to say that the recovered image depends on the choice of diffusion-encoding gradient. A function $E(\mathbf{q})$ independent of T1 and T2 relaxation effects and quantifying signal attenuation due solely to diffusion can be defined

$$E(\mathbf{q}) = S(\mathbf{q})/S_0, \tag{3.32}$$

where $S_0$ is the MR signal obtained in the absence of diffusion encoding.

One of the most significant results of the theory of diffusion imaging is the fact that $E(\mathbf{q})$ is related to the EAP, denoted $\overline{P}(\mathbf{r})$, via Fourier transformation [25] (derivation provided in Appendix A)

$$\overline{P}(\mathbf{r}) = \int_{\mathbb{R}^3} E(\mathbf{q})\, e^{-\imath 2\pi\,(\mathbf{q}\cdot\mathbf{r})} d\mathbf{q}. \tag{3.33}$$

Hence, by measuring $E(\mathbf{q})$ at a sufficiently large set of points on a Cartesian grid in $\mathbf{q}$-space, it is possible to recover the EAP at each voxel by means of the fast Fourier transform. Initially, phase corruptions caused by biological motion limited this method to the imaging of inanimate objects. Later modifications using the modulus Fourier transform in lieu of the complex Fourier transform solved these problems and lead to the technique of diffusion spectral imaging (DSI) [90, 95]. It goes without saying that the major limitation of DSI is its acquisition requirements, requiring roughly 500 DW-images corresponding to points distributed throughout $\mathbf{q}$-space. Unfortunately, such acquisition requirements put DSI far beyond the scope of practical applications of DW-MRI.

### 3.2.4 High angular resolution diffusion imaging

As a compromise between the relatively short acquisition times of DTI and the multimodal diffusion profiles available to DSI, the method of high angular resolution diffusion imaging (HARDI) has been proposed [6, 31, 40, 92, 93]. In terms of acquisition, HARDI is similar to DTI except that it requires measuring a considerably higher number of DW-images, typically between 80 and 100. The associated encoding gradients are required to satisfy

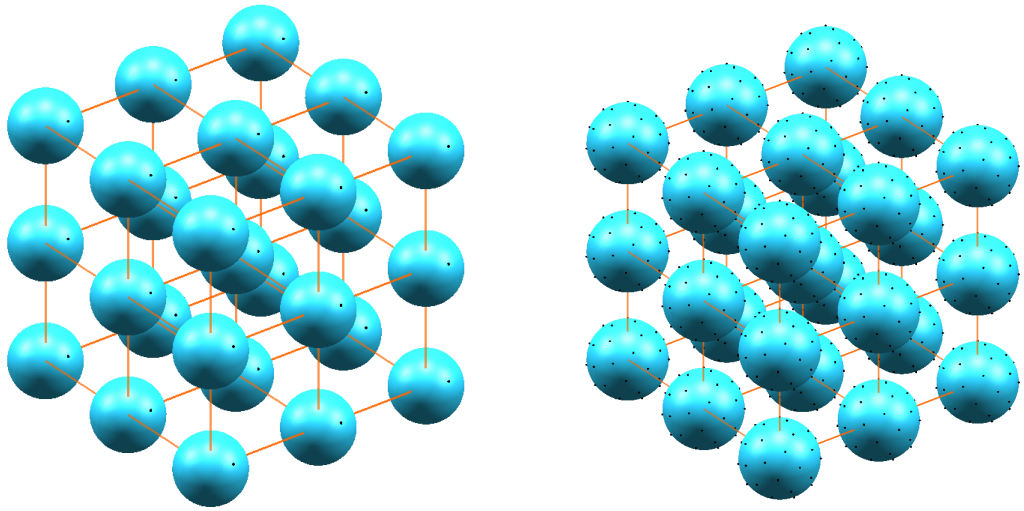
a constant b-value, changing only the orientation $\mathbf{u}$ between acquisitions. This allows the HARDI signal recovered from K space to be expressed as a function of spatial location $\mathbf{r}$ and diffusion-encoding gradient orientation $\mathbf{u}$

$$S(\mathbf{r}, \mathbf{u}) = S_0(\mathbf{r})\, e^{-bD(\mathbf{r},\mathbf{u})}. \tag{3.34}$$

The value at a specific spatial location is commonly rewritten as a spherical function in the notation

$$S(\mathbf{u}) = S_0\, e^{-bD(\mathbf{u})}, \tag{3.35}$$

where $D(\mathbf{u})$ is the orientation-dependent ADC. Note that the HARDI ADC is a general spherical function in contrast to the DTI ADC, which was restricted to the shape of an ellipse. Mathematically, the HARDI signal in (3.34) is viewed as a scalar function on the $\mathbb{R}^3 \times \mathbb{S}^2$ manifold, formally written $S(\mathbf{r}, \mathbf{u}) : \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$. This function is measured at sample locations by acquiring DW-images. A DW-image can be acquired using an encoding gradient at any spatial orientation $\mathbf{u} \in \mathbb{S}^2$, but every DW-image is recovered from K space using the fast Fourier transform and is therefore discrete in the spatial domain $\mathbb{R}^3$. This discretization of the $\mathbb{R}^3 \times \mathbb{S}^2$ manifold in the spatial domain can be visualized as a discrete lattice in $\mathbb{R}^3$ where at each lattice point is a sphere representing the possible $\mathbf{u} \in \mathbb{S}^2$ sample locations (the blue spheres in Figure 3.6). Acquiring a DW-image with a specific encoding gradient orientation $\mathbf{u}_j$ is equivalent to simultaneously sampling every sphere in the lattice at orientation $\mathbf{u}_j$ (the black points in Figure 3.6(a)). Therefore, a collection of DW-images acquired with different gradient orientations has corresponding sample locations which can be visualized as a discrete lattice of discrete locations on the unit sphere (the black points in Figure 3.6(b)).



(a) A single DW-image (b) A collection of DW-images

Figure 3.6: Sample locations in HARDI signal domain

The increased number of samples available in HARDI allows the use of diffusion models which are more accurate and realistic than the DTI tensor model. Both parametric and non-parametric approaches have been proposed to facilitate the computation of a number of HARDI processing techniques, including the analysis of the ADC [4, 31, 40], orientation distribution function (ODF) [32, 65, 91, 93], fibre orientation distribution function (FODF) [6,88], and persistent angular structure (PAS) [50] to name just a few. All of these methods will benefit from the denoising technique proposed in this thesis. However, analysis of the filter will mainly be performed using the ODF.

### Q-Ball imaging and orientational distribution function

The constraint of imposing a constant b-value during HARDI acquisition has the related effect of restricting the values of $\mathbf{q}$ to belong to a spherical shell in $\mathbf{q}$-space. HARDI, therefore, reduces the DSI burden of sampling an entire three-dimensional Cartesian grid while, at the same time, maximizes the amount of orientational information being measured in $\mathbf{q}$-space for a given number of samples. Q-Ball Imaging (QBI), developed by Tuch, is a technique which uses the $\mathbf{q}$-space measurements to extract orientational information from the EAP. More precisely, QBI approximates the orientational distribution function (ODF)

which is defined as the radial projection of the EAP along a given orientation $\mathbf{u}$

$$\psi(\mathbf{u}) = \frac{1}{Z} \int_0^\infty \overline{P}(R\mathbf{u}) dR, \tag{3.36}$$

where Z is a normalization constant and $\overline{P}$ is the EAP. Thus, while $\overline{P}$ quantifies the likelihood with which a spin is displaced by $\mathbf{r}$, $\psi(\mathbf{u})$ quantifies the likelihood with which the spin is expected to move in direction $\mathbf{u}$. For this reason, QBI works on the assumption that the orientational structure of the tissue is of primary interest.

The backbone of the QBI ODF approximation lies in the Funk-Radon Transform (FRT). The FRT evaluated at $\mathbf{u}$ is defined as the integral of a spherical function along the great circle with pole $\mathbf{u}$ (note that the great circle is defined to be the intersection of the unit sphere with a plane which has normal $\mathbf{u}$ and goes through the origin). Tuch has shown that the ODF can be approximated by applying the FRT to the attenuation signal $E(\mathbf{q})$ [91,93]

$$\psi(\mathbf{u}) \approx \mathcal{R}\{E\}(\mathbf{u}) = \frac{1}{Z} \iiint_{|\mathbf{v}|=1} \delta(\mathbf{u} \cdot \mathbf{v}) E(\mathbf{v}) d\mathbf{v}. \tag{3.37}$$

Succinctly, integrating the attenuation signal along a great circle obtains an approximation of the probability of diffusion in the direction of the great circle's pole. To discuss the accuracy of this approximation, consider evaluating the FRT in the unit $\hat{z}$ direction. In this case, the exact relationship in cylindrical coordinates between $\overline{P}$ and $\mathcal{R}\{E\}(\mathbf{e}_z)$ was shown to be equal to:

$$\mathcal{R}\{E\}(\mathbf{e}_z) = 2\pi q_0 \int_{-\infty}^{\infty} \int_0^{2\pi} \int_0^{\infty} \overline{P}(\rho, \phi, z) J_0(2\pi q_0 \rho) d\rho d\phi dz, \tag{3.38}$$

where $q_0$ is the radius of the shell in $\mathbf{q}$-space and $J_0$ is a Bessel function of the first kind and order 0. The degree of error in this ODF approximation depends on how well the Bessel kernel $J_0$ is localized in $\rho$ (i.e.. if $J_0$ was the Dirac delta function, then we would obtain the exact radial projection along $\mathbf{e}_z$). The main lobe of the Bessel kernel has a width which depends on the value of $q_0$. A larger encoding gradient amplitude results in a greater $q_0$ and a narrower Bessel kernel. A narrow Bessel kernel produces less blurring and greater angular resolution in the ODF approximation. Unfortunately, the large gradient amplitude (and associated b-value) required for accurate Q-ball imaging significantly attenuates signal levels. Fortunately, the resulting trade-off between angular resolution and SNR can be improved through the filtering of HARDI signal noise - the topic of this thesis.

A number of authors have pointed out that the radial projection computed in (3.36) is not a true probability distribution [1, 89]. Since the expression for the Ensemble Average Propagator in spherical coordinates can be written

$$\iiint_{\mathbb{R}^3} \overline{P}(\mathbf{r})d\mathbf{r} = \int_0^{\pi}\int_0^{2\pi}\left(\int_0^{\infty}\overline{P}(R\mathbf{u})R^2dR\right)\sin(\theta)d\phi d\theta, \tag{3.39}$$

the ODF expression would have to contain an extra $R^2$ term in order to be a true probability density obtained by marginalizing out the radial variable. Therefore, a modified diffusion descriptor known as the orientational probability distribution function (OPDF) has been proposed

$$\Phi(\mathbf{u}) = \int_0^{\infty}\overline{P}(R\mathbf{u})R^2dR, \tag{3.40}$$

along with a number of methods to approximate its value. For example, [89] proposes to use an FRT approximation similar to that of (3.37), except that $E$ (the inverse Fourier transform of $\overline{P}$) is replaced by the Laplacian of $E$ (the inverse Fourier transform of $\overline{P}R^2$)

$$\Phi(\mathbf{u}) \approx \mathcal{R}\{\Delta E\}(\mathbf{u}). \tag{3.41}$$

A difficulty in computing this approximation is that the Laplacian is composed of derivatives in all coordinate directions while $E$ is only measured in the inclination and azimuth directions. Therefore, it is necessary to make an assumption about the behaviour of $E$ in the radial direction in order to approximate the derivative of $E$ with respect to a radial variable. [89] overcomes this difficulty using the model of (3.34) with the additional assumption of a constant ADC. It was shown that OPDF approximations lead to an increased ability to characterize complex microstructures.

## 3.3 HARDI noise model

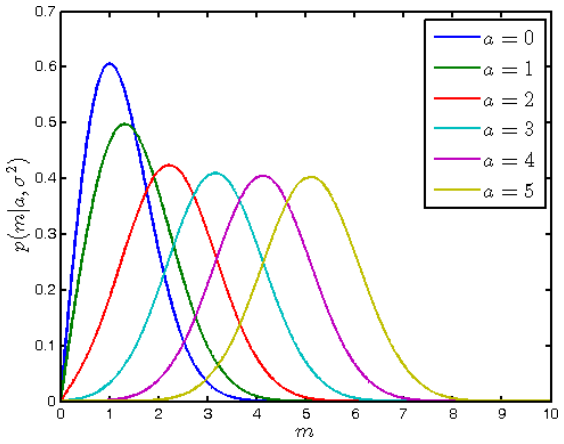
The MR image is acquired in the Fourier domain (i.e., the $k$-space), followed by the procedures of frequency demodulation and inverse transformation. The result is a complex-valued image, whose magnitude is typically the displayed output image. In the case that the frequency-domain data is contaminated by zero-mean AWG noise, the complex amplitude $M$ of the noisy observation $Ae^{\imath\theta} + N$, with $N = N_r + \imath N_i$ , is given by (see (3.23))

$$M = \sqrt{(A\cos\theta + N_r)^2 + (A\sin\theta + N_i)^2}, \tag{3.42}$$

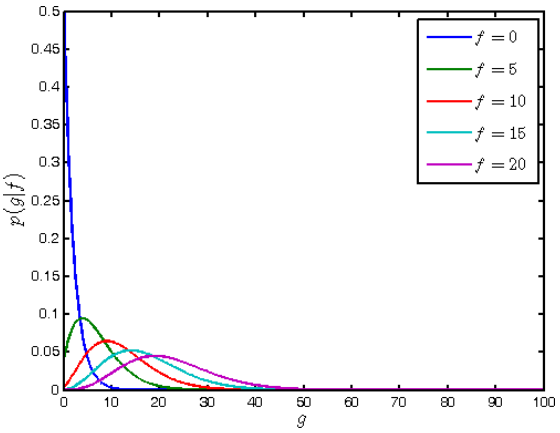
where $A$ stands for the true image amplitude, while $N_r$ and $N_i$ are mutually independent AWG noises of standard deviation $\sigma$, and $\theta \in [0, 2\pi)$ is an arbitrary phase shift. In this case, $M$ can be shown to follow the Rician conditional distribution model that is given by [45, 58]

$$p_{M|A}(m|a) = \begin{cases} \frac{m}{\sigma^2} \exp\left\{-\frac{a^2+m^2}{2\sigma^2}\right\} I_0\left(\frac{am}{\sigma^2}\right), & m \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.43}$$

where $I_0$ denotes the $0^{th}$-order modified Bessel function of the first kind. Figure 3.7(a) depicts several typical graphs of $p_{M|A}(m|a)$ corresponding to a range of the values of $A$ and $\sigma = 1$. As can be seen from the figure, for $A > 3\sigma$, the Rician probability density is closely approximated by a Gaussian distribution [45]. However, for the lower values of $A$, the density becomes more asymmetric and noticeably heavy-tailed. Specifically, for $A = 0$, $M$ follows a Rayleigh distribution model.




(a) Rician pdf's corresponding to different values of $A$ and $\sigma = 1$ in (3.43)




(b) Non-central chi square distribution corresponding to different values of $F$ in (3.46)

Figure 3.7: MRI-related statistical distributions

Rician noise renders impractical a straightforward application of many filtering strategies which attempt to infer the true value $A$ from an estimation of $\mathcal{E}\{M\}$, the expected value of $M$. The difficulties lie in the complex relation between $\mathcal{E}\{M\}$ and $A$

$$\mathcal{E}\{M\} = \sigma\sqrt{\pi/2}\, L_{1/2}(-A^2/2\sigma^2), \tag{3.44}$$

where $L_v(x)$ denotes a Laguerre polynomial which, for $v = 1/2$, is given by

$$L_{1/2}(x) = e^{x/2}\left[(1-x)I_0\left(\frac{-x}{2}\right) - xI_1\left(\frac{-x}{2}\right)\right]. \tag{3.45}$$

At the same time, a normalized version $G = (M/\sigma)^2$ of the squared magnitude $M^2$ can be shown to be distributed according to the non-central chi square (NCCS) distribution with parameter $F = (A/\sigma)^2$ whose conditional density is given by

$$p_{G|F}(g|f) = \begin{cases} \frac{1}{2}\exp\left\{-\frac{g+f}{2}\right\} I_0(\sqrt{fg}), & g \geq 0 \\ 0, & \text{otherwise,} \end{cases} \tag{3.46}$$

where $f \in \mathbb{R}^+$. Figure 3.7(b) shows a number of typical graphs of $p_{G|F}$ corresponding to a set of different values of $F$. A better understanding of this figure can be obtained from the fact that (3.42) suggests

$$G = F + 2\sqrt{F}\xi + \eta, \tag{3.47}$$

where $\xi = (N_r \cos\theta + N_i \sin\theta)/\sigma$ and $\eta = (N_r^2 + N_i^2)/\sigma^2$. Thus $G$ can be viewed as a noisy version of $F$, where the noise has both additive and multiplicative components. Specifically, it should be noted that $\xi$ obeys a normal distribution with zero mean and unit variance, while $\eta$ follows an exponential distribution with its mean and variance equal to 2 and 4, respectively. Moreover, the expectation of $G$ now has a very simple relation to $F$, which is given by

$$\mathcal{E}\{G\} = F + 2. \tag{3.48}$$

It is the simplicity of (3.48) which has been a principal impetus for the development of various MRI denoising methods, which have been applied to the squared magnitude $G$, rather than to its original value $M$. In some manner, all these methods aim at recovering a close approximation of the average value $\mathcal{E}\{G\}$, followed by the estimation of $F$ through the subtraction of the bias of 2. Note that, once an estimate of $F$ has been obtained, its associated amplitude $A$ in (3.42) can be recovered through simple square root and re-normalization.

## 3.4 Spherical harmonics

Spherical harmonics (SH) are the natural way to decompose functions on the unit sphere $\mathbb{S}^2$, similar to the Fourier transformation for functions in Euclidean space. Recalling that it is common to view the HARDI signal as a collection of spherical functions, it is not surprising that SHs have become a backbone tool for analyzing HARDI signals [4, 32, 40]. This thesis will make use of the SH coefficients as a method to compare spherical function similarity. Therefore, in this section, we introduce some of the basic SH properties. As a general reference on the subject, the reader is referred to [44].

Defined by the SH differential equation, SHs are the eigenfunctions of the Laplace-Beltrami operator $\Delta_b Y_n^l = -l(l+1)Y_n^l$, where $n$ is called the *order* and $l$ is the *phase factor*. The solution is given explicitly by

$$Y_n^l(\theta,\phi) = \sqrt{\frac{(2n+1)(n-l)!}{4\pi(n+l)!}} P_n^l(\cos\theta) e^{il\phi}, \tag{3.49}$$

where $P_n^l$ are associated Legendre polynomials. A subset of these functions are shown in Figure 3.8. As is evident from both the notation $Y_n^l$ and the figure, SHs are divided into *bands*. All functions in the same band have the same order and each band contains $2n+1$ functions distinguished by their phase factor.

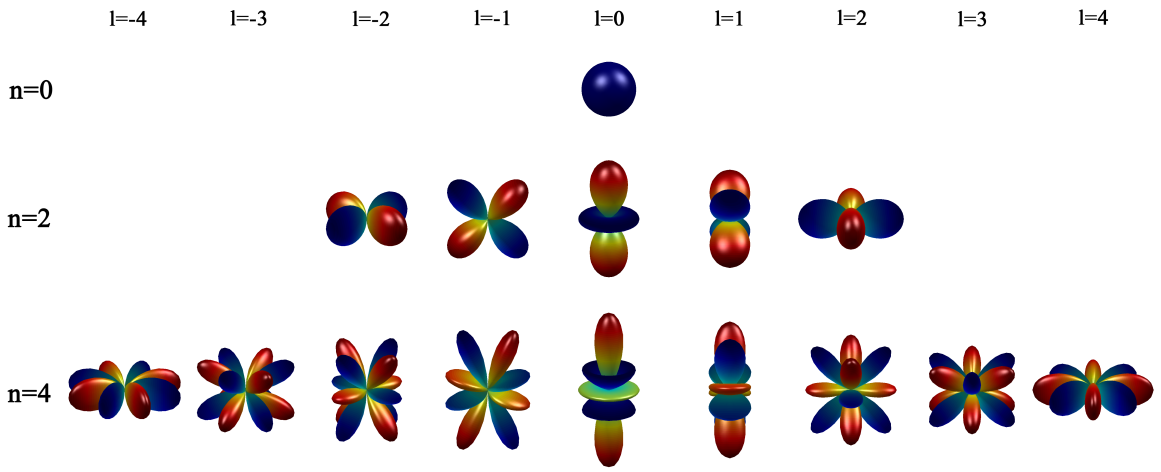



Figure 3.8: Real spherical harmonic basis functions (even orders up to and including n=4)

There are a number of important SH properties relevant to the development of the proposed method. First, SHs form an orthonormal basis for functions on the unit sphere. Consequently, the energy of a spherical signal is related to its spectral coefficients by Parseval's theorem. It is therefore possible to divide a spherical function's total signal energy into band energies based on band coefficients. Moreover, a rotated SH can be written as a linear combination of all non-rotated SHs of the same order. As a consequence, if a signal is in $\text{span}\{Y_n^l\}_{|l|\leq n}$ for a given order $n$, then it's rotated version is also in $\text{span}\{Y_n^l\}_{|l|\leq n}$ (since the rotated signal can be written as a linear combination of rotated SHs, and each rotated SH can itself be written as a linear combination of non-rotated SHs). Noting that rotation does not change signal energy, we conclude that the band energies are invariant to

signal rotations. The second property we point out will only be mentioned in Chapter 6's discussion concerning future work. It can be seen that the zero phase ($l = 0$) SHs are zonal functions (middle column of Fig 3.8). Therefore, zero phase coefficients also provide a signal description which is invariant to azimuth rotations. Finally, we note that SHs of even order are antipodal symmetric whereas SHs of odd order are antipodal asymmetric. Therefore, antipodal symmetric functions are represented using only even order basis functions. This is particularly useful for HARDI signal representation, since diffusion measurement is a symmetric process.

# Chapter 4

# Proposed Methodology

This chapter introduces the proposed non-local means (NLM) approach to filtering the HARDI signal. The NLM method estimates a sample intensity's true value by computing a weighted average of other similar samples contained in the image. In contrast to linear filtering where averaging weights are fixed, NLM defines the weights adaptively based on a neighbourhood similarity measure. Adaptive weights are particularly useful in DW-MRI since the spatial dependency of diffusivity in brain tissue makes it impossible to find a single set of linear filter parameters which is optimal for all types of diffusion signals. As opposed to similar methods in the field, the proposed NLM filtering includes samples from the entire $\mathbb{R}^3 \times \mathbb{S}^2$ manifold in the weighted average, defines neighbourhoods based solely on a sample's local orientational diffusion characteristics, and uses weights designed to be invariant to both spatial rotations as well as to the particular sampling scheme in use.

The chapter is organized as follows. In Section 4.1 we review the original NLM filter, the essence of which is neighbourhood comparison. Since a neighbourhood is nothing more than a group of neighbouring samples, it is appropriate to investigate *sample similarity measures* as a precursor to *neighbourhood similarity measures*. Section 4.2 investigates possible sample similarity measures specific to MRI statistics and introduces the proposed correlation similarity measures. It should be noted that the proposed sample similarity measures are not only useful in developing a novel NLM method to filter HARDI data, but can also improve traditional NLM filtering techniques applied to Euclidean MR images. As discussed in Section 4.3, the natural neighbourhood for HARDI diffusion signals is a subset of $\mathbb{S}^2$ and the proposed NLM method, despite averaging samples from the entire $\mathbb{R}^3 \times \mathbb{S}^2$ manifold, conducts neighbourhood comparisons in the $\mathbb{S}^2$ domain. Section 4.3 proposes two novel neighbourhood similarity measures for $\mathbb{S}^2$. Finally, Section 4.4 discusses the removal of a bias connected to MRI noise statistics.

## 4.1 Non-local means filter

The concept of NLM filtering was first proposed in [20] for the case of zero-mean additive white Gaussian (AWG) noise contamination of images in $\mathbb{R}^2$. Let $X$ and $Y$ denote the original and observed images, respectively, and $X_s$ and $Y_s$ denote the intensities corresponding to image pixel $s \in I \equiv \{0, 1, \ldots, N-1\}$. Then, for zero-mean AWG noise, $X_s$ is equal to the expected value $\mathcal{E}\{Y_s\}$. Many filtering techniques assume $Y$ to be an ergodic process, in which case $\mathcal{E}\{Y_s\}$ can be estimated by a spatial average of $\{Y_t\}_{t \in J_s}$, where $J_s \subseteq I$ denotes the set of pixel indices with associated intensities distributed identically to $Y_s$. In most practical situations, it is impossible to deduce from noisy data the sets of samples which realize an identical underlying distribution. It is therefore sensible to compute $\hat{X}_s = \hat{\mathcal{E}}\{Y_s\}$, the estimate of $X_s = \mathcal{E}\{Y_s\}$, by including every image pixel in a weighted average:

$$\hat{X}_s = \hat{\mathcal{E}}\{Y_s\} = \frac{1}{C_s} \sum_{t \in I} w_{s,t}\, Y_t, \qquad \text{with } C_s = \sum_{t \in I} w_{s,t}, \tag{4.1}$$

where $w_{s,t} \geq 0$ quantifies the "contribution" of a target pixel $t \in I$ to the estimate of the source pixel $s$. Ideally, the weights $w_{s,t}$ should reflect the degree of similarity between the source $X_s$ and target $X_t$. The Gaussian range filter, for example, defines the weights as $w_{s,t} = \mathrm{sim}(Y_s, Y_t) = \exp(-\frac{1}{2}\frac{|Y_s - Y_t|^2}{h})$. However, the approximation of expected value is degraded by the random nature of $Y_s$ and $Y_t$ which frequently allows the range filter to have a low-valued $w_{s,t}$ even when underlying $X_s$ and $X_t$ are similar (and vice versa). NLM, on the other hand, defines the weights in a more robust manner, based on the comparison of neighbourhood similarity

$$w_{s,t} = \mathrm{sim}(Y_{N_s}, Y_{N_t}) \quad \text{with } Y_{N_s} = \{Y_k\}_{k \in N_s},\ Y_{N_t} = \{Y_k\}_{k \in N_t} \text{ and } N_s, N_t \subseteq I, \tag{4.2}$$

where $N_s$ and $N_t$ denote the set of pixel indices with associated pixels in the neighbourhood of $s$ and $t$, respectively (see Figure 4.1). The effectiveness of this non-local processing method is due to the self-similarity apparent in natural images [19,23]. Thus, for example, the original choice of neighbourhood comparison in [20] was defined to be

$$w_{s,t} = \exp\left\{ -\frac{1}{h} \sum_{k \in \Omega} \beta_k \, |Y_{s-k} - Y_{t-k}|^2 \right\}, \tag{4.3}$$

where $\Omega$ is the index set representing a symmetric neighbourhood of the centre-of-image coordinates. For example, in an image with two spatial coordinates $(x, y)$, $\Omega$ could be defined as $\Omega = \{|x| \leq L_x, |y| \leq L_y\}$ for some positive integers $L_x$ and $L_y$. Then $Y_{N_s} = \{Y_{s-k}\}_{k \in \Omega}$

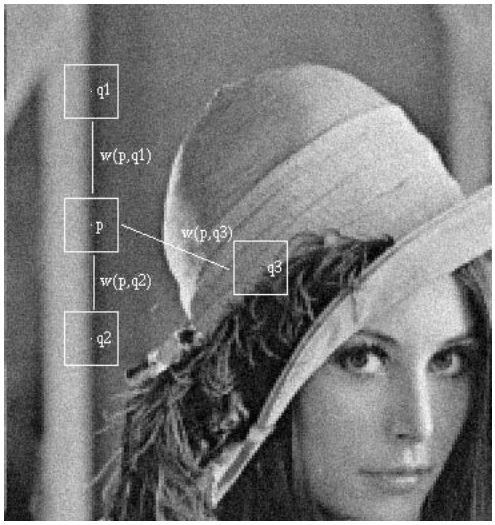



Figure 4.1: The NLM approach. The picture and notation are from the original NLM paper [20]. Source pixel p is estimated by an average of all possible target pixels. Only three of the possible target pixels are highlighted and labelled q1, q2, and q3. Similar pixel neighbourhoods give large weights - w(p,q1),w(p,q2) - while dissimilar neighbourhoods give small weights - w(p,q3).

and $Y_{N_t} = \{Y_{t-k}\}_{k\in\Omega}$. The "fine-tuning" parameters $\{\beta_k\}_{k\in\Omega}$ in (4.3) are intended to weight the domain of summation and they should be chosen to satisfy $\sum_{k\in\Omega}\beta_k = 1$, while $h > 0$ controls the overall amount of smoothing imposed by the filter. Specifically, higher values of $h$ tend to result in overly smoothed output images, whereas lower values produce rather mild filtering effect. As a general rule, an optimal value of $h$ should be chosen adaptively according to the level of noise in $Y$.

To facilitate our considerations, we note that the NLM weights in (4.3) admit an alternative definition as given by [30, 85]

$$w_{s,t} = \prod_{k\in\Omega} (\mathrm{SNL}_{s,t,k})^{\frac{\beta_k}{h}}, \tag{4.4}$$

with $\mathrm{SNL}_{s,t,k}$ being a sample similarity measure defined as

$$\mathrm{SNL}_{s,t,k} = \exp\left\{-|Y_{s-k} - Y_{t-k}|^2\right\}. \tag{4.5}$$

Thus, NLM compares each pixel in the source neighbourhood to a corresponding pixel in the target neighbourhood, and computes the neighbourhood similarity by taking the

product of the resulting sample similarity values. It is important to note that the value of $\mathrm{SNL}_{s,t,k}$ is always bounded between 0 and 1, and it can be shown (see [30, 55] for more details) that its choice in (4.4) and (4.5)is optimal in the case of AWG noise contamination.

## 4.2 Sample similarity measures for MRI

The neighbourhood comparison provided in (4.4) conducts a sample-to-sample comparison between corresponding source and target neighbourhood pixels. While the similarity measure in (4.5) is optimal for AWG contamination, there exist a number of alternative statistical approaches to compare samples in the presence of non-Gaussian noise. These statistical methods are described next for the MRI-specific cases of Rician and non-central chi squared (NCCS) statistics followed by the introduction of two novel correlation similarity measures. The proposed sample similarity measures have applications to NLM filtering in $\mathbb{R}^2$ (classic MR images) using (4.4) in (4.1) and to NLM filtering on $\mathbb{R}^3 \times \mathbb{S}^2$ (HARDI images) using (4.29) and (4.30) in (4.24).

### 4.2.1 Statistical approaches

In [30], it was suggested to set the similarity measure $\mathrm{SNL}_{s,t,k}$ to be equal to the posterior probability of $X_{s-k} = X_{t-k}$ conditioned on observations of $Y_{s-k}$ and $Y_{t-k}$. Formally,

$$\mathrm{SNL}_{s,t,k} = P(X_{s-k} = X_{t-k} | Y_{s-k}, Y_{t-k}), \tag{4.6}$$

which, for the case $\beta_k/h = 1, \forall k \in \Omega$, leads to following definition of the NLM weights

$$w_{s,t} = \prod_{k \in \Omega} P(X_{s-k} = X_{t-k} | Y_{s-k}, Y_{t-k}). \tag{4.7}$$

It is worthwhile noting that, under the assumption of statistical independence of the intensities of the original image $X$, the weights in (4.7) can be viewed as the posterior probability of the neighbourhoods $\{X_{s-k}\}_{k\in\Omega}$ and $\{X_{t-k}\}_{k\in\Omega}$ to consist of the same intensities, conditioned on observation of their corresponding noisy values $\{Y_{s-k}\}_{k\in\Omega}$ and $\{Y_{t-k}\}_{k\in\Omega}$, respectively [30]. Although the assumption of statistical independence is an obvious oversimplification, it is often employed in NLM filtering to render the final estimation scheme computationally feasible.

The similarity measure in (4.6) seems to have a serious theoretical flaw in the case of continuous random variables $X_{s-k}$ and $X_{t-k}$, in which case it is always equal to zero [43,

p. 111]. To overcome this difficulty, it was suggested in [85] to introduce an auxiliary random variable $U_k \equiv X_{s-k} - X_{t-k}$ and set $\mathrm{SNL}_{s,t,k}$ to the value of the conditional density $p_{U_k|Y_{s-k},Y_{t-k}}(u_k|y_{s-k}, y_{t-k})$ at $u_k = 0$. Alternatively, one can use a different auxiliary variable $V_k = X_{s-k}/X_{t-k}$, and set $\mathrm{SNL}_{s,t,k}$ to be equal to the value of $p_{V_k|Y_{s,k},Y_{t,k}}(v_k|y_{s-k}, y_{t-k})$ at $v_k = 1$ [85]. For the convenience of referencing, the above similarity measures will be referred to below as the *subtractive* and the *rational* similarity measures, respectively. Note that although these measures are alike in their underlying philosophy, they lead to substantially different denoising schemes, as detailed below.

In the present work, we explore both the subtractive and rational similarity measures for two different types of input data, namely for the measured magnitude image $M$ which follows a Rician distribution (see 3.43) and its squared normalized version $G$ which follows a non-central chi square distribution (NCCS) (see 3.46). The main contribution of the next propositions is to provide closed-form expressions for the similarity measures which result from using a subtractive $U_k$ on $G$ images, and a rational $V_k$ on $M$ images. Unfortunately, we have been unable to derive closed-form expressions for the remaining two combinations (*viz.*, "rational" $G$ and "subtractive" $M$). In these cases, the measures need to be computed numerically – an approach which should be avoided in practice due to its high computational cost.

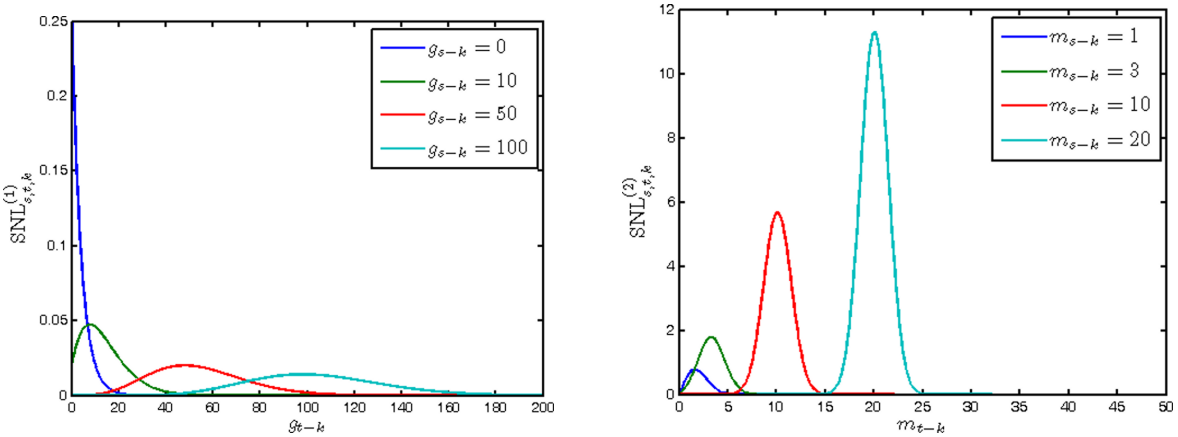
**Proposition 4.2.1.** *Let $G = (M/\sigma)^2$ be a squared and normalized version of the MR magnitude image $M$. Moreover, let $F = (A/\sigma)^2$, where $A$ denotes the true signal amplitude. Then, the subtractive similarity measure for the associated non-central chi square distribution $\mathrm{SNL}^{(1)}_{s,t,k}$ is given by*[1]

$$\begin{aligned}\mathrm{SNL}^{(1)}_{s,t,k} &= p_{F_{s-k}-F_{t-k}|G_{s-k},G_{t-k}}(0 \mid g_{s-k}, g_{t-k}) \\ &= \frac{1}{4}\exp\left\{-\frac{g_{s-k}+g_{t-k}}{4}\right\} I_0\left(\frac{\sqrt{g_{s-k}\,g_{t-k}}}{2}\right). \end{aligned} \tag{4.8}$$

The proof of Proposition 4.2.1 is provided in Appendix B.1, while Fig. 4.2(a) shows four different $\mathrm{SNL}^{(1)}_{s,t,k}$ output curves when the source input variable $g_{s-k}$ is fixed and the target input variable $g_{t-k}$ is allowed to vary.

**Proposition 4.2.2.** *Let $M$ be the MR magnitude image. Moreover, let $A$ be the original signal amplitude. Then, the rational similarity measure for the associated Rician distribu-*

[1]Here and hereafter, we use the standard statistical formalism for denoting random variables and their associated realizations by capital letters and their lower-case counterparts, respectively.




(a) Subtractive similarity measure $\mathrm{SNL}^{(1)}_{s,t,k}$ (b) Rational similarity measure $\mathrm{SNL}^{(2)}_{s,t,k}$

Figure 4.2: Subtractive and rational similarity measures

*tion* $\mathrm{SNL}^{(2)}_{s,t,k}$ *is given by*

$$\begin{aligned}\mathrm{SNL}^{(2)}_{s,t,k} &= p_{A_{s-k}/A_{t-k}|M_{s-k},M_{t-k}}(1 \mid m_{s-k}, m_{t-k}) \\ &= \frac{m_{s-k}\, m_{t-k}}{\sigma^2} \exp\left\{-\frac{m^2_{s-k} + m^2_{t-k}}{4\sigma^2}\right\} I_0\left(\frac{m_{s-k}\, m_{t-k}}{\sigma^2}\right). \end{aligned} \tag{4.9}$$

The proof of Proposition 4.2.2 can be found in Appendix B.2. Four $\mathrm{SNL}^{(2)}_{s,t,k}$ output curves are shown in Fig. 4.2(b), where $\sigma$ is set to be equal to 1, source input variable $m_{s-k}$ is fixed, and target input variable $m_{t-k}$ is allowed to vary.

It can be seen from Figure 4.2 that $\mathrm{SNL}^{(1)}_{s,t,k}$ and $\mathrm{SNL}^{(2)}_{s,t,k}$ attractively appear to reflect the main properties of their corresponding noise distributions, nevertheless, both similarity measures share a number of critical drawbacks. An intuitive way to understand these drawbacks is through the definition of a distance measure $d_{s,t,k}$ between two intensities $Y_{s-k}$ and $Y_{t-k}$ as given by

$$d_{s,t,k} = -\log(\mathrm{SNL}_{s,t,k}). \tag{4.10}$$

For example, in the case of AWG noise, $d_{s,t,k} = (Y_{s-k} - Y_{t-k})^2$. Naturally, $\mathrm{SNL}_{s,t,k}$ can be recovered back from its associated $d_{s,t,k}$ using

$$\mathrm{SNL}_{s,t,k} = e^{-d_{s,t,k}}. \tag{4.11}$$

Unfortunately, in the case of $\mathrm{SNL}^{(1)}_{s,t,k}$ and $\mathrm{SNL}^{(2)}_{s,t,k}$, the distance measure $d_{s,t,k}$ has a number of undesirable properties, which are highlighted in the points below.

1. Neither $\mathrm{SNL}^{(1)}_{s,t,k}$ nor $\mathrm{SNL}^{(2)}_{s,t,k}$ attain their maximum value when the source and target input arguments are equal. Particularly, for a fixed value of $g_{s-k}$ (resp. $m_{s-k}$), the $\mathrm{SNL}^{(1)}_{s,t,k}$ measure (resp. the $\mathrm{SNL}^{(2)}_{s,t,k}$ measure) is maximal at some $g_{t-k} < g_{s-k}$ (resp. $m_{t-k} < m_{s-k}$). In terms of $d_{s,t,k}$, for a fixed value of $g_{s-k}$ (resp. $m_{s-k}$), the minimum distance is achieved at some $g_{t-k} < g_{s-k}$ (resp. $m_{t-k} < m_{s-k}$).

2. The maximal values that $\mathrm{SNL}^{(1)}_{s,t,k}$ and $\mathrm{SNL}^{(2)}_{s,t,k}$ can attain depend on the values of the source input $g_{s-k}$ and $m_{s-k}$, respectively (both plots in Figure 4.2 have four curves which attain different maximum values). From a purely applicational point of view, this fact suggests that the measures are not scale invariant, and as a result, the weights $w_{s,t}$ in (4.1) are defined *not only* by how similar compared intensities are, but also by their absolute values.

3. As can be seen from Figure 4.2(b), the distance $d_{s,t,k}$ can be negative for high values of $m_{s-k}$, which is an unacceptable property of $d_{s,t,k}$ as a metric.

In light of these deficiencies, it is evident that neither $\mathrm{SNL}^{(1)}_{s,t,k}$ nor $\mathrm{SNL}^{(2)}_{s,t,k}$ is adequate for measuring sample similarity in the cases of Rician or NCCS noise. In the next section, we propose a new similarity measure which is free of the aforementioned limitations.

### 4.2.2 Proposed correlation sample similarity measures

In this section, we modify the subtractive similarity measure for non-central chi squared noise and the rational similarity measure for Rician noise to produce two novel "correlation" similarity measures.

#### Correlation subtractive similarity measure for NCCS noise

To derive the proposed similarity measure in a consistent and intuitive manner, we start with the definition of the subtractive similarity measure (SSM) for an arbitrary noise distribution [85]

$$\begin{aligned}\mathrm{SSM} &= p_{F_{s-k}-F_{t-k}|G_{s-k},G_{t-k}}(0 \mid g_{s-k}, g_{t-k}) \\ &= \int_0^\infty p_{F_{s-k}|G_{s-k},G_{t-k}}(f \mid g_{s-k}, g_{t-k})\, p_{F_{t-k}|G_{s-k},G_{t-k}}(f|g_{s-k}, g_{t-k})\, df. \end{aligned} \tag{4.12}$$

Alternatively, using Bayes' theorem one can show that the SSM is proportional to, *viz.* [30, 85]

$$\text{SSM} \propto \int_0^\infty p_{G_{s-k}|F_{s-k}}(g_{s-k} \mid f)\, p_{G_{t-k}|F_{t-k}}(g_{t-k} \mid f)\, df. \tag{4.13}$$

In moving from (4.12) to (4.13), it has been assumed that the prior probability $p_{F_{z-k}}$ (where $z \in \{s, t\}$ and $k \in \Omega$) is uniform. Moreover, it has also been assumed that the original intensity at a particular position is conditionally independent of the noisy intensity at a different position, given the noisy intensity at the particular position. Formally,

$$p_{F_{z_1-k}|G_{z_1-k},G_{z_2-k}} = p_{F_{z_1-k}|G_{z_1-k}}. \tag{4.14}$$

To overcome the limitations of $\text{SNL}^{(1)}_{s,t,k}$ and $\text{SNL}^{(2)}_{s,t,k}$ as discussed in the previous section, we interpret the right-hand side of (4.13) as an inner product between the *likelihood functions* $L_{g_{s-k}}(\cdot)$ and $L_{g_{t-k}}(\cdot)$, with $L_g(f)$ given by

$$L_g(f) = p_{G|F}(g \mid f). \tag{4.15}$$

Note that, in general, the likelihood functions $L_{g_{s-k}}(\cdot)$ and $L_{g_{t-k}}(\cdot)$ have unequal norms, and as a result their inner product is not maximized when $g_{s-k} = g_{t-k}$ (which would be a natural and desirable property for a sample similarity measure to have). To overcome this shortcoming, we suggest to normalize the inner product, thereby converting it into a correlation subtractive similarity measure (CSSM) according to

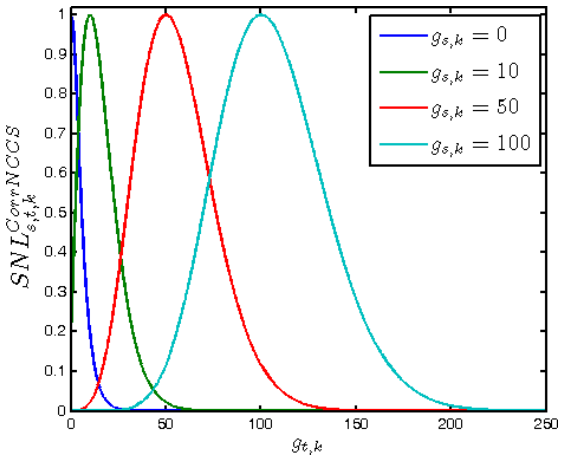
$$\text{CSSM}_{s,t,k} = \frac{\langle L_{g_{s-k}}, L_{g_{t-k}} \rangle}{\|L_{g_{s-k}}\|_2 \, \|L_{g_{t-k}}\|_2}, \tag{4.16}$$

where $\langle x, y \rangle = \int_0^\infty x(f) y(f)\, df$ and $\|x\|_2 = \sqrt{\langle x, x \rangle}$. $\text{CSSM}_{s,t,k}$ is free of all the major limitations of $\text{SNL}^{(1)}_{s,t,k}$ and $\text{SNL}^{(2)}_{s,t,k}$. In particular, $\text{CSSM}_{s,t,k}$ is always smaller or equal to 1, and it achieves a consistent maximum value of 1 whenever $g_{s-k} = g_{t-k}$.

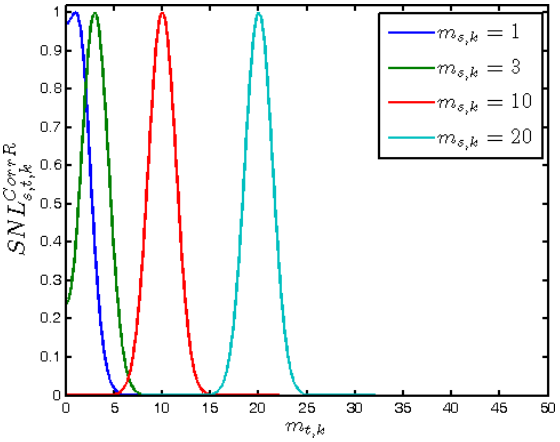
For the case of NCCS noise, the CSSM measure can be shown to have a neat closed-form expression which is given by

$$\text{SNL}^{(3)}_{s,t,k} = \frac{I_0\left(\sqrt{\hat{g}_{s-k}\, \hat{g}_{t-k}}\right)}{\sqrt{I_0(\hat{g}_{s-k})\, I_0(\hat{g}_{t-k})}}, \quad \text{where } \hat{g} = g/2. \tag{4.17}$$

Multiple curves of $\text{SNL}^{(3)}_{s,t,k}$ are shown in Figure 4.3(a), where each curve is drawn with a fixed source input $g_{s-k}$ and varying target input $g_{t-k}$. It can be seen from the graphs that the shape of each curve is similar to those in Figure 4.2(a). However, unlike the plots in Figure 4.2(a), each curve in Figure 4.3(a) is maximized when $g_{t-k} = g_{s-k}$ and the maximum value is always equal to 1.




(a) Proposed CSSM $\mathrm{SNL}^{(3)}_{s,t,k}$ for the case of NCCS noise distribution




(b) Proposed CRSM $\mathrm{SNL}^{(4)}_{s,t,k}$ for the case of Rician noise distribution

Figure 4.3: Proposed correlation similarity measures

**Correlation rational similarity measure for Rician noise**

To derive an expression for a correlation rational similarity measure (CRSM) in the case of Rician noise, we first state that the rational similarity measure (RSM) is given by [30]

$$\begin{aligned}\mathrm{RSM} &= p_{A_{s-k}/A_{t-k}|M_{s-k},M_{t-k}}(1 \mid m_{s-k}, m_{t-k}) \\ &= \int_0^\infty a\, p_{A_{s-k}|M_{s-k},M_{t-k}}(a \mid m_{s-k}, m_{t-k})\, p_{A_{t-k}|M_{s-k},M_{t-k}}(a \mid m_{s-k}, m_{t-k})\, da. \end{aligned} \tag{4.18}$$

Similar to the SSM case, one can show that the RSM is proportional to, *viz.* [30, 85]

$$\mathrm{RSM} \propto \int_0^\infty p_{M_{s-k}|A_{s-k}}(m_{s-k} \mid a)\, p_{M_{t-k}|A_{t-k}}(m_{t-k} \mid a)\, a da. \tag{4.19}$$

and this integral can be interpreted as a weighted inner product $\langle x, y\rangle_a = \int_0^\infty x(a)y(a)\, a\, da$, where $a\, da$ can be viewed as a "modified" integration measure. Using this notation, we define the correlation rational similarity measure (CRSM) according to

$$\mathrm{CRSM}_{s,t,k} = \frac{\left\langle L_{m_{s-k}}, L_{m_{t-k}} \right\rangle_a}{\|L_{m_{s-k}}\|_2 \, \|L_{m_{t-k}}\|_2}. \tag{4.20}$$

The CRSM for Rician noise is given by

$$\mathrm{SNL}^{(4)}_{s,t,k} = \frac{I_0\left(\frac{m_{s-k}m_{t-k}}{2\sigma^2}\right)}{\sqrt{I_0\left(\frac{m^2_{s-k}}{2\sigma^2}\right) I_0\left(\frac{m^2_{t-k}}{2\sigma^2}\right)}}. \tag{4.21}$$

A number of $\mathrm{SNL}^{(4)}_{s,t,k}$ curves are shown in Fig. 4.3(b), , where each curve is drawn with a fixed source input $m_{s-k}$ and varying target input $m_{t-k}$. Once again, one can observe that the curves are similar in shape to those in Fig. 4.2(b). However, unlike Fig. 4.2(b), each $\mathrm{SNL}^{(4)}_{s,t,k}$ curve is maximized when $m_{t-k} = m_{s-k}$ and the maximum value is always equal to 1.

There are two important facts about $\mathrm{SNL}^{(3)}_{s,t,k}$ and $\mathrm{SNL}^{(4)}_{s,t,k}$ that deserve to be paid special attention. In particular:

1. The values of $\mathrm{SNL}^{(3)}_{s,t,k}$ and $\mathrm{SNL}^{(4)}_{s,t,k}$ (as given by (4.17) and (4.21), respectively) are equal under the substitution $g_{z-k} = (m_{z-k}/\sigma)^2$, which is precisely the relation between the magnitude image $M$ and its squared normalized version $G$. This fact suggests that, when using $\mathrm{SNL}^{(3)}_{s,t,k}$ on $G$ and $\mathrm{SNL}^{(4)}_{s,t,k}$ on $M$, two neighbourhoods of an original MR image and their corresponding squared and normalized versions are equally similar.

2. Using the fact that, for sufficiently large $x$, it holds that

$$I_0(x) \approx \frac{e^x}{\sqrt{2\pi x}}, \tag{4.22}$$

and substituting (4.22) into (4.21) instead of the original Bessel functions, we obtain

$$\mathrm{SNL}^{(4)}_{s,t,k} \approx \exp\left\{-\frac{|m_{s-k} - m_{t-k}|^2}{4\sigma^2}\right\}. \tag{4.23}$$

The above approximation holds with a high precision for relatively large values of SNR (i.e., for $\frac{m_k}{\sigma}, \frac{m_l}{\sigma} \gg 1$). This is an exceptional property of $\mathrm{SNL}^{(4)}_{s,t,k}$, since it is known that Rician noise in MRI converges to Gaussian noise when SNR increases. Likewise, the proposed $\mathrm{SNL}^{(4)}_{s,t,k}$ measure converges to the form of (4.5), whose optimality for Gaussian noises was proven in [30].

## 4.3 Extending NLM to the $\mathbb{R}^3 \times \mathbb{S}^2$ manifold

With suitable sample similarity measures in hand, we are now in a position to extend NLM to the $\mathbb{R}^3 \times \mathbb{S}^2$ manifold. Mathematically, one can view the HARDI signal $S_n$ corresponding to voxel $n \in I_n \equiv \{0, 1, \ldots, N-1\}$ as a scalar function defined on the unit sphere $\mathbb{S}^2$, i.e. $S_n(\mathbf{u}) : \mathbb{S}^2 \to \mathbb{R}$. The set of all spherical functions in the imaging volume, $\{S_n\}_{n \in I_n}$, is evaluated along a predefined number of spatial orientations $\{\mathbf{u}_k\}_{k \in I_k}$ where $I_k \equiv \{0, 1, \ldots, K-1\}$ (recall Fig. 3.6). Let $Y_n(\mathbf{u}_k)$ denote the noise-contaminated observation of intensity $X_n(\mathbf{u}_k)$ corresponding sample $k$ from spherical function $n$. Then, using samples from the entire $\mathbb{R}^3 \times \mathbb{S}^2$ manifold, an estimate $\hat{\mathcal{E}}\{Y_{s_n}(\mathbf{u}_{s_k})\}$ of $\mathcal{E}\{Y_{s_n}(\mathbf{u}_{s_k})\}$ can be computed

$$\hat{\mathcal{E}}\{Y_{s_n}(\mathbf{u}_{s_k})\} = \frac{1}{C_{s_n,s_k}} \sum_{t_n \in I_n} \sum_{t_k \in I_k} \omega_{s_n,s_k,t_n,t_k} Y_{t_n}(\mathbf{u}_{t_k}), \quad \text{with } C_{s_n,s_k} = \sum_{t_n \in I_n} \sum_{t_k \in I_k} \omega_{s_n,s_k,t_n,t_k}. \tag{4.24}$$

The NLM weights $\omega_{s_n,s_k,t_n,t_k}$ reflect the similarity in structure between source and target neighbourhoods and in order for an NLM approach to be successful in HARDI, the HARDI signal must be self-similar with respect to the neighbourhood chosen. HARDI signals imaging neural fibres composing white matter (Figure 4.4) admit a number of possible definitions of the neighbourhood, and the choice has an impact on what surrounding characteristics must be similar in order to credit two samples as similar. In the present discussion we will consider a spatial neighbourhood in $\mathbb{R}^2$ (only two dimensions are considered for simplicity) and a spherical neighbourhood in $\mathbb{S}^2$.

Self-similarity of the HARDI signal with respect to spatial neighbourhoods in $\mathbb{R}^2$ requires the subject's neural fibres to have similarly shaped paths. In order for NLM to be successful, neural fibres should have repetitive spatial structure. Consider two straight fibres travelling parallel in the $\hat{x}$ direction (HARDI signal shown in Figure 4.5(a)), then the similarity in shape between these two fibres will be replicated in each DW-image (for example, Figure 4.5(b) shows the replicated spatial structure in the DW-MR image at orientation $\mathbf{u} = [1\,0\,0]^T$). Consider estimating a particular sample in the $\mathbf{u} = [1\,0\,0]^T$ DW-MR image at spatial location $(x, y) = (4, 2)$ (see the magenta marker in Figure 4.5(b)). Every sample in Figure 4.5(b) corresponding to one of the two fibres has a similar spatial neighbourhood (for example, note the similarity between the neighbourhoods of sample $(x, y) = (4, 2)$, shown in magenta, and sample $(x, y) = (2, 4)$, shown in blue). This is the motivation behind applying classical NLM filtering to each DW-image independently. However, the search for similar target samples need not be restricted to one DW-image alone. For instance, in the absence of noise, the $\mathbf{u} = [-1\,0\,0]^T$ DW-image should be identical to

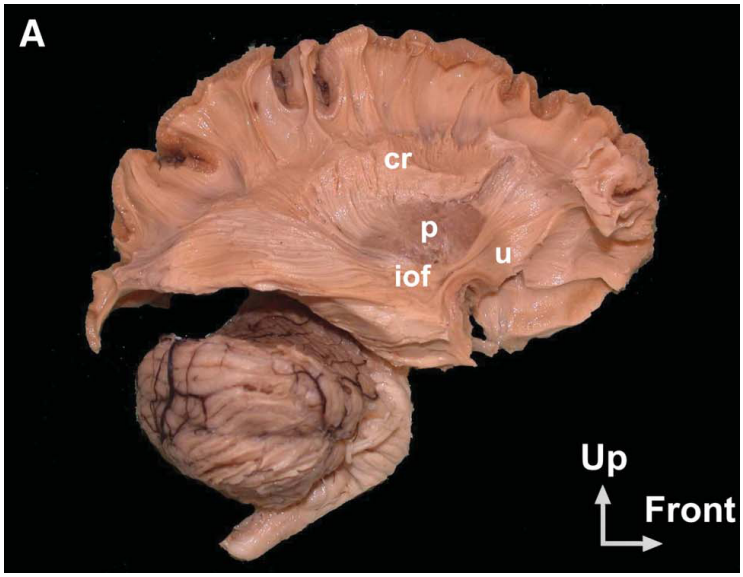
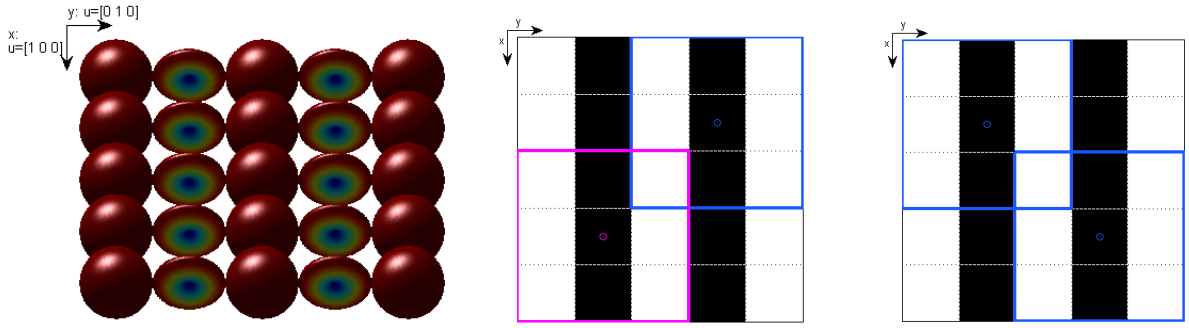



Figure 4.4: Neural fibres in the human brain [77]

the $\mathbf{u} = [1\,0\,0]^T$ image (Figure 4.5(c)) due to the symmetry in diffusion measurements. Therefore, allowing the search for similar target neighbourhoods to be performed over all DW images may increase the number of similar samples included in the NLM weighted average.

Unfortunately, spatial neighbourhood comparison is unable to detect numerous similar samples when fibres follow more complicated paths. Consider the HARDI signal of a U-shaped fibre shown in Figure 4.6(a) and suppose we wish to estimate the sample located at orientation $\mathbf{u} = [1\,0\,0]^T$ and spatial location $(x, y) = (4, 2)$ (see the magenta marker in Figure 4.6(c)). Beyond the similar target samples that can be found in the $\mathbf{u} = [-1\,0\,0]^T$ DW-image, we note that since the fibre curves into the $\hat{y}$ direction, there is also a similar target sample in the $\mathbf{u} = [0\,1\,0]^T$ image at $(x, y) = (2, 3)$ (Figure 4.6(d)). However, this sample has a spatial neighbourhood which differs significantly from the source neighbourhood due to the shape of the underlying fibre. Consequently, although similar, this sample will be given a low weight in the NLM average. For similar reasons, the vector NLM method of [99] described in Section 2.6 is also unable to properly weight similar vectors in images containing curving fibres due to spatial dissimilarity. The problem is that curving neural fibres have limited repetitions in spatial structure. However, the similar samples excluded by spatial neighbourhoods can be given appropriate weights in the NLM average by comparing neighbourhoods in $\mathbb{S}^2$.



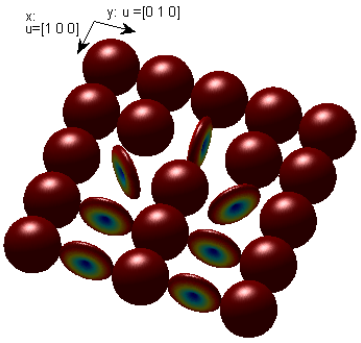

(a) HARDI signal for two parallel fibres (b) DW-image in the $\mathbf{u} = [1\,0\,0]^T$ direction (c) DW-image in the $\mathbf{u} = [-1\,0\,0]^T$ direction

Figure 4.5: Similar spatial neighbourhoods in multiple DW images

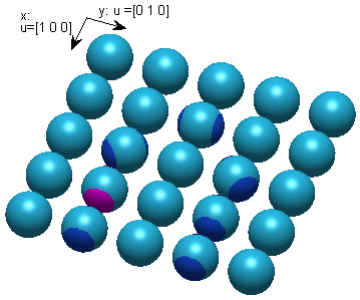
Self-similarity of the HARDI signal with respect to neighbourhoods in $\mathbb{S}^2$ requires repetitive structure of localized orientational diffusion characteristics. Thus, antipodal sample locations on the same spherical function will always have similar local $\mathbb{S}^2$ neighbourhoods (due to diffusion measurement symmetry). Furthermore, spherical functions located at the spatial position of a fibre crossing can have four or more orientations with similar $\mathbb{S}^2$ neighbourhoods. The search for target samples with similar $\mathbb{S}^2$ neighbourhoods need not be restricted to the source spherical function; the search can be extended to include target spherical functions at all spatial locations. Since neural fibres possess consistent diffusion properties (consistent microstructure) throughout the spatial dimensions, the $\mathbb{S}^2$ neighbourhood comparison is able to detect similar samples even as the fibre curves through space. Returning to the U-shaped fibre example with source sample at $\mathbf{u} = [1\,0\,0]^T$ and $(x, y) = (4, 2)$, Figure 4.6(b) illustrates the source neighbourhood location in magenta. Comparing Figure 4.6(a) and Figure 4.6(b), it can be seen that the blue neighbourhoods in Figure 4.6(b) represent the locations of similar target neighbourhoods. Thus, it is proposed that HARDI signals will have a greater degree of self-similarity using neighbourhoods which exclude spatial dimensions and are defined completely in $\mathbb{S}^2$.

Exploring NLM filtering of HARDI signals using neighbourhoods defined on $\mathbb{S}^2$ is the main contribution of this thesis. For notational simplicity, and without loss of generality, in defining the neighbourhood similarity measure we will restrict target samples to belong to the source sphere $Y_{s_n}$. Essentially, we will solve the problem of extending NLM to $\mathbb{S}^2$. Suppressing the $s_n$ and $t_n$ subscripts, we rewrite Equation 4.24 as

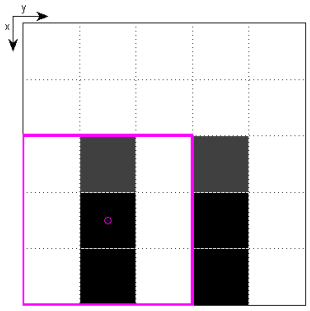
$$\hat{\mathcal{E}}\{Y(\mathbf{u}_s)\} = \frac{1}{C_s} \sum_{t \in I_k} \omega_{s,t} Y(\mathbf{u}_t), \qquad \text{with } C_s = \sum_{t \in I_k} \omega_{s,t}. \tag{4.25}$$



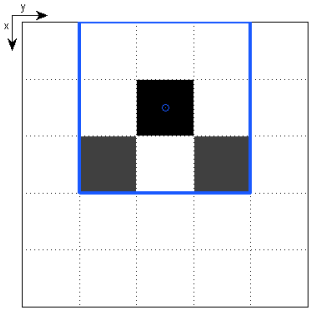

(a) HARDI signal of a U-shaped fibre




(b) Locations of similar target $\mathbb{S}^2$ neighbourhoods




(c) Spatial source neighbourhood from $\mathbf{u} = [1\,0\,0]^T$ DW-image




(d) Dissimilar spatial target neighbourhood from $\mathbf{u} = [0\,1\,0]^T$ DW-image

Figure 4.6: Neighbourhoods of complex fibres

Note that (4.25) is trivially extended to NLM filtering of HARDI data on $\mathbb{R}^3 \times \mathbb{S}^2$ by permitting additional target spherical functions to contribute samples into the average.

### 4.3.1 $\mathbb{S}^2$ neighbourhood similarity measure

Defining neighbourhood comparison on $\mathbb{S}^2$ is not trivial. To describe the challenges and proposed methods, we first define an appropriate neighbourhood for the spherical HARDI signal. Let $N_s$ and $N_t$ be two spherical neighbourhoods corresponding to spatial orienta-

tions $\mathbf{u}_s$ and $\mathbf{u}_t$, respectively, which are defined as

$$N_s = \left\{ u \in \mathbb{S}^2 \middle| \ \cos^{-1}(|u \cdot u_s|) \leq \eta \right\} \tag{4.26}$$

$$N_t = \left\{ u \in \mathbb{S}^2 \middle| \ \cos^{-1}(|u \cdot u_t|) \leq \eta \right\} \tag{4.27}$$

where $\eta \leq \pi/2$ controls the size of the neighbourhood. An important characteristic of the neighbourhoods defined in Equation 4.26 is antipodal symmetry, a result of the absolute value taken inside of the arc cosine. This definition of the neighbourhood has been chosen in connection with the antipodal symmetry of the diffusion signal. However, other applications requiring the filtering of non-symmetric spherical functions can safely discard the absolute value.

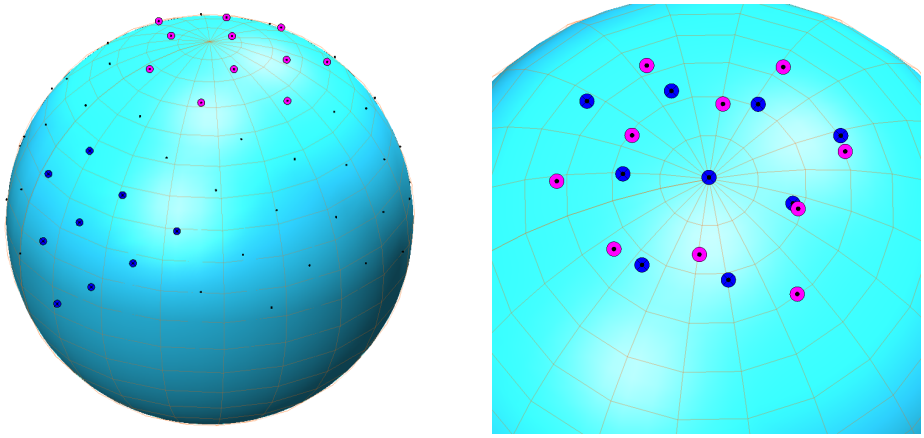
A necessary step of neighbourhood comparison is the alignment of their centres (for instance, in $\mathbb{R}^2$ two square neighbourhoods are simply translated to overlap and subsequently compared pixelwise). In $\mathbb{S}^2$, we propose to align neighbourhoods by rotating both centres to $\hat{z}$. However, a number of difficulties arise. The first obstacle is that, due to azimuth rotations, there are infinitely many ways to align the two neighbourhoods (in fact, this is the reason why the sphere $\mathbb{S}^2$ is equivalent to the quotient space $SO(3)/SO(2)$), making it difficult to justify one method of alignment over another. A second obstacle occurs in attempting to align sampled neighbourhoods for sample-to-sample comparison. Only a limited number of restrictive sampling schemes are uniform and rotation invariant [104]. Therefore, in general, it is not possible to perfectly overlap two sampled neighbourhoods (see Figure 4.7). In fact, two neighbourhoods with the same characteristic $\eta$ often do not even contain the same number of samples. We propose two neighbourhood comparison methods to overcome the above difficulties.

### Sample-to-neighbourhood similarity measure

The inability to overlap neighbourhood samples prevents the direct comparison of a sample from the source neighbourhood with a single corresponding sample from the target neighbourhood. However, one can instead compare the source sample to every sample in the target neighbourhood (Figure 4.8), weighting each comparison according to proximity. Let the $\mathbb{S}^2$ metric be the standard metric of arclength, then the distance between $\mathbf{u}_l \in N_s$ and $\mathbf{u}_j \in N_t$ is computed

$$d(\mathbf{u}_l, \mathbf{u}_j) = \cos^{-1}(|\mathbf{R}_s \mathbf{u}_l (\mathbf{R}_t \mathbf{u}_j)^T|), \tag{4.28}$$

where $\mathbf{R}_s, \mathbf{R}_t \in \mathrm{SO}(3)$ are appropriate rotation matrices aligning source and target neighbourhoods together in the $\hat{z}$ direction.



(a) Two sampled neighbourhoods with $\eta = \pi/6$ (antipodal points excluded for simplicity)
Patch 1 (magenta, ◯)
Patch 2 (blue, ⨁)

(b) Non-overlapping samples after neighbourhood alignment

Figure 4.7: $\mathbb{S}^2$ neighbourhood alignment

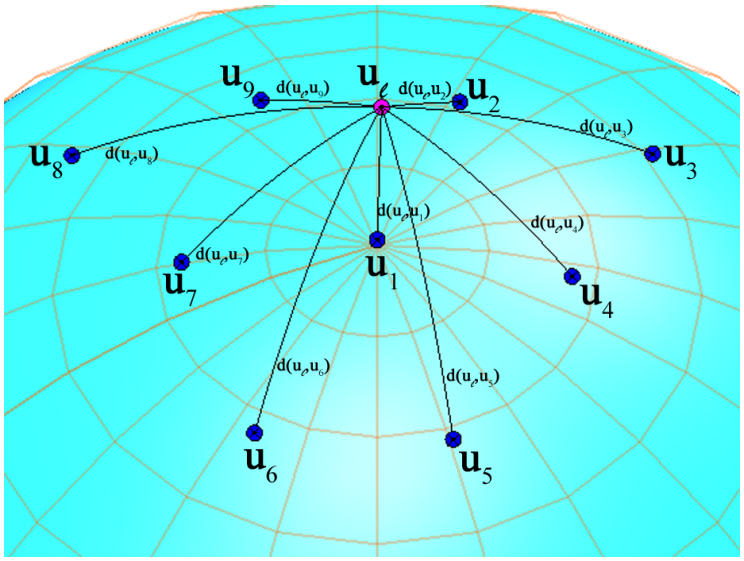



Figure 4.8: $\mathrm{SNL}_{u_l,N_t}$ is computed by comparing a source sample $\mathbf{u}_l$ to every target sample in $N_t$ (antipodal points are excluded from the figure for simplicity)

We define the sample-to-neighbourhood similarity measure between a source sample $\mathbf{u}_l \in N_s$ and the entire target neighbourhood $N_t$

$$\mathrm{SNL}_{u_l,N_t} = \frac{1}{C_l} \sum_{u_j \in N_t} \exp\left\{-\frac{d(\mathbf{u}_l, \mathbf{u}_j)}{\alpha^2}\right\} (\mathrm{SNL}_{u_l,u_j})^{\frac{\beta_k}{h}}, \quad \text{with } C_l = \sum_{u_j \in N_t} \exp\left\{-\frac{d(\mathbf{u}_l, \mathbf{u}_j)}{\alpha^2}\right\}. \tag{4.29}$$

For HARDI denoising, $\mathrm{SNL}_{u_l,u_j}$ is one of the MRI sample similarity measures proposed in Section 4.2.2. Note that target samples near $\mathbf{u}_l$ have a greater impact on the sum. Sensitivity to target sample proximity is controlled by the tuning parameter $\alpha$.

To complete the neighbourhood comparison, each source pixel is compared to the entire target neighbourhood and the NLM weights are defined as

$$w_{s,t} = \prod_{u_l \in N_s} \mathrm{SNL}_{u_l,N_t}. \tag{4.30}$$

Furthermore, this approach can be made independent to the method of neighbourhood alignment (i.e.. the choice of $\mathbf{R}_s$ and $\mathbf{R}_t$) by simply changing the distance metric to depend only on the sample's elevation angle

$$d(\mathbf{u}_l, \mathbf{u}_j) = |\cos^{-1}(|\mathbf{R}_s \mathbf{u}_l \mathbf{e}_z^T|) - \cos^{-1}(|\mathbf{R}_t \mathbf{u}_j \mathbf{e}_z^T|)|, \tag{4.31}$$

where $\mathbf{e}_z$ is the unit vector in the $\hat{z}$ direction.

Needless to say, the sample-to-neighbourhood similarity measure is extremely computationally intensive. For example, if neighbourhood comparison in $\mathbb{R}^2$ requires $N$ sample comparisons, then the current proposed scheme requires $N^2$ sample comparisons. Unfortunately, the complexity of HARDI data further amplifies this algorithm's computational requirements. Therefore, in the following section we will propose a suboptimal neighbourhood similarity measure which is more computationally efficient.

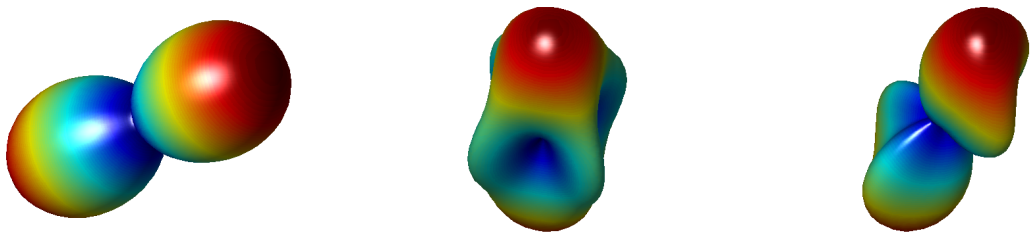
### RILES neighbourhood similarity measure

In the previous section, the use of direct sample comparisons lead to a neighbourhood similarity measure which is currently computationally infeasible. In this section, we propose a different method of comparing neighbourhoods using the transform domain of spherical harmonics. Toward this end, we define $W_{\mathbf{v}}(\mathbf{u})$ to be a smooth and symmetric window function supported around the antipodal directions $\mathbf{v}$ and $-\mathbf{v}$, with $\mathbf{v} \in \mathbb{S}^2$. Then, multiplication of a HARDI signal $S(\mathbf{u})$ by $W_{\mathbf{v}}(\mathbf{u})$ localizes the former around $\mathbf{v}$ and $-\mathbf{v}$, effectively

extracting a symmetric neighbourhood $\{S(\mathbf{u})\}_{\mathbf{u}\in N_v}$. The symmetric Gauss-Weierstrass window defines such a function (see Figure 4.9):

$$W_{\mathbf{v}}(\mathbf{u}) = \sum_{n=0,2,4,\ldots} \frac{2n+1}{4\pi} e^{-\rho n(n+1)} P_n(\mathbf{u}\cdot\mathbf{v}), \tag{4.32}$$

where $P_n(\mathbf{u}\cdot\mathbf{v})$ are Legendre Polynomials. The Gauss-Weierstrass kernel, constructed using spherical harmonics (SH), is designed to have similar properties to the Gaussian kernel in Euclidean space [24,26]. In the above formula, the summation over even SH orders ensures that the window is symmetric. The parameter $\rho$ is used to control the effective support of the kernel around the antipodal centres $\mathbf{v}$ and $-\mathbf{v}$. Note that, being the product of two symmetric functions, the windowed function $S(\mathbf{u})W_{\mathbf{v}}(\mathbf{u})$ is also symmetric.



(a) Kernel $W_{\mathbf{v}}(\mathbf{u})$ (b) HARDI signal $S_n(\mathbf{u})$ (c) Windowed signal $W_{\mathbf{v}}(\mathbf{u})S_n(\mathbf{u})$

Figure 4.9: Gauss-Weierstrass windowing

Decomposing the windowed signal into a composition of the SH basis functions provides coefficients indicative to the behaviour of $S_n(\mathbf{u})$ in the neighbourhood of $\mathbf{v}$. We write the SH coefficients as

$$\{C_{n,l}\}_{\substack{n=0,2,4,\ldots,N_{max}\\ |l|\leq n}}, \tag{4.33}$$

where $N_{max}$ is the maximum SH order inclusive. Due to symmetry, odd order coefficients of $S_n(\mathbf{u})W_{\mathbf{v}}(\mathbf{u})$ are zero.

In general, SH coefficients are not rotation invariant (a rotated windowed function $S(\mathbf{Ru})W_{\mathbf{v}}(\mathbf{Ru})$, $\mathbf{R}\in SO(3)$ will, in general, have coefficients different from those of $S(\mathbf{u})W_{\mathbf{v}}(\mathbf{u})$). However, the $\ell^2$ energy of SH coefficients belonging to a single band is invariant under arbitrary rotations. Therefore, a rotation invariant description of the windowed signal is given by the vector

$$\overline{f}_{\mathbf{v}} = \begin{bmatrix} f(0) \\ f(2) \\ \vdots \\ f(N_{max}) \end{bmatrix}, \tag{4.34}$$

where each element in the vector is the $\ell^2$ energy of the even n$^{th}$ order coefficients

$$f(n) = \frac{\sum_{\ell=-n}^{n} |c_{n,\ell}|^2}{2n+1}. \tag{4.35}$$

We refer to $\overline{f}_{\mathbf{v}}$ as the rotation invariant localized energy signature (RILES) vector. The distance between two neighbourhoods $N_s$ and $N_t$ is then defined to be the Euclidean distance $\left\|\overline{f}_{\mathbf{u}_s} - \overline{f}_{\mathbf{u}_t}\right\|_2^2$. Finally, the RILES neighbourhood similarity measure and corresponding weights are defined

$$w_{s,t} = \exp\left\{-\frac{\left\|\overline{f}_{\mathbf{u}_s} - \overline{f}_{\mathbf{u}_t}\right\|_2^2}{h^2}\right\}, \tag{4.36}$$

where $h > 0$ is a tuning parameter controlling the overall amount of smoothing imposed by the filter. Specifically, higher values of $h$ tend to result in overly smoothed output images, whereas smaller values produce rather mild filtering effects.

### 4.3.2 Adding spatial dimensions to $\mathbb{S}^2$ neighbourhood similarity measures

In some diffusion MRI applications, such as DTI and HARDI compressed sensing, a minimal number of samples are taken in the $\mathbb{S}^2$ domain. In these cases, there may not be enough information in the $\mathbb{S}^2$ domain to make an accurate comparison between spherical neighbourhoods. Adding a spatial dimension to the neighbourhood will increase the amount of information being compared making the weight computation more robust. Although the detection of similar samples in complex fibre shapes will be forfeited, it is a necessary sacrifice for an accurate NLM average. Therefore, in these situations, we propose a spatial neighbourhood of spherical neighbourhoods (see Figure 4.10). Source and target neighbourhoods are overlapped and corresponding spherical neighbourhoods are compared. Thus, instead of determining if the source and target samples admit similar local diffusion processes, this comparison determines if all local diffusion processes happening in the spatial vicinity of source and target samples are similar. This method is similar to the one proposed in [99], however, it differs in two significant ways. Unlike [99], the proposed neighbourhood is localized in $\mathbb{S}^2$. Furthermore, the proposed method allows target samples with an orientation different from the source sample to contribute to the approximation of the source sample. An alternative method would be to continue using neighbourhoods in

$\mathbb{S}^2$ and instead multiply each neighbourhood comparison by a term based on the geometric distance between source and target samples in the $\mathbb{R}^3 \times \mathbb{S}^2$ manifold. This approach would be similar to the bilateral filter commonly used in image processing.

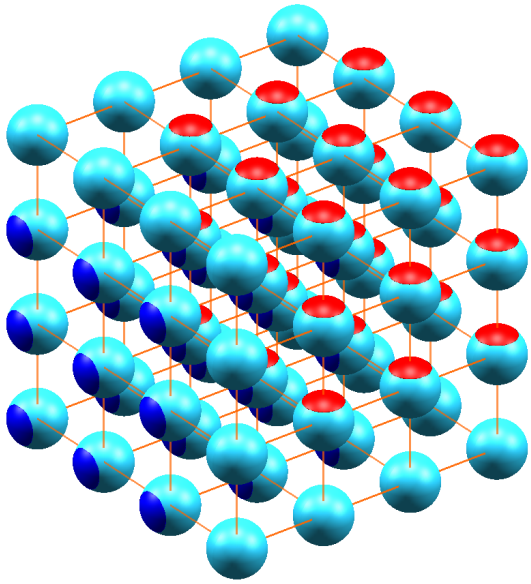


Figure 4.10: Two neighbourhoods containing both $\mathbb{R}^3$ and $\mathbb{S}^2$components

## 4.4 Bias removal

The HARDI NLM weighted average of (4.24) provides an estimate of a sample's expected value $\hat{\mathcal{E}}\{Y_{s_n}(\mathbf{u}_{s_k})\}$. In the case of Rician noise, the relation between the original signal amplitude $A$ and the noisy measurement $Y$ is not trivial. However, if NLM averaging is applied to the squared normalized HARDI image $G$, then one can take advantage of the relation in (3.48) and use $\hat{\mathcal{E}}\{G_{s_n}(\mathbf{u}_{s_k})\}$ to estimate the related original amplitude $A = \sigma\sqrt{\mathcal{E}\{G\} - 2}$ as

$$A_{s_n}(\mathbf{u}_{s_k}) \approx \sigma \left[ \max\left\{ \hat{\mathcal{E}}\{G_{s_n}(\mathbf{u}_{s_k})\} - 2,\, 0 \right\} \right]^{1/2}, \tag{4.37}$$

where the $\max(\cdot)$ operator is used to avoid complex estimates. Note that, even though applying the $\max(\cdot)$ operator may seem like a purely ad-hoc procedure, it has been shown to be optimal in the ML sense in [80]. This estimate has also been used previously in [100] in the context of non-local means.

# Chapter 5

# Results

## 5.1 Reference methods

The performance of the proposed NLMS2 and RILES methods has been compared with several standard algorithms on both simulated and real HARDI data. As the first reference method, Gaussian kernel smoothing (referred to hereafter as GKS) of Section 2.3 was performed on the squared normalized signal followed by the bias removal procedure specified in Section 4.4. GKS was applied to spherical functions with differing smoothness in order to compare with and demonstrate the adaptivity of the proposed NLM weights. As the second reference method, the total variation (TV) method of Section 2.5 has been used[1]. In what follows, this method is referred to as vector TV (vTV). The final reference method used for numerical comparison was the NLM filter of Section 2.6, referred to below as NLMv. In the case of all reference methods under comparison, their respective parameters have been set based on the guidelines specified in their associated papers.

Two new approaches to NLM filtering of HARDI images are proposed in this thesis. Specifically, the first approach compares $\mathbb{S}^2$ neighbourhoods using a sample-to-neighbourhood similarity measure. For the convenience of referencing, this filtering approach is referred below to as NLMS2 (with "S2" referring to the $\mathbb{S}^2$ neighbourhood). The second approach compares $\mathbb{S}^2$ neighbourhoods in the transform domain of spherical harmonics and is referred to below as RILES. All the acronyms of the proposed and reference algorithms are summarized in Table 5.1.

[1]The method was implemented using code kindly provided by the author

Table 5.1: Acronyms of the proposed and reference algorithms

| Algorithm name | Reference | Input image type |
|---|---|---|
| GKS | [26] | G |
| vTV | [56] | M |
| NLMv | [99] | G |
| NLMS2 | Proposed | G |
| RILES | Proposed | G |

## 5.2 Simulated data

### 5.2.1 Data generation

The HARDI signal (3.35) was simulated at each spatial location using a generalization of DTI known as the *multi-tensor model*. The multi-tensor model considers the EAP to be a mixture of $n$ Gaussians corresponding to $n$ crossing fibres. Consequently, the HARDI signal at a specific spatial location is modelled

$$S(\mathbf{u}) = S_0 \sum_{j=1}^{n} a_j e^{-b\mathbf{u}^T \mathbf{D}_j \mathbf{u}}, \tag{5.1}$$

where $\mathbf{D}_j$ is the diffusion tensor associated with the $j^{\text{th}}$ fibre and $a_j$ are proportionality constants satisfying $\sum_{j=1}^{n} a_k = 1$. A given noise-free intensity from the HARDI signal $A \in S$ was corrupted by Rician noise to give a magnitude intensity of

$$M = \sqrt{(A + n_r)^2 + n_i^2}, \tag{5.2}$$

where $n_r$, $n_i \sim \mathcal{N}(0, \sigma^2)$ are independent Gaussian random variables, which are also assumed to be independent across the image domain. The standard deviation $\sigma$ was set to attain a desired SNR using the relation $\text{SNR} = \bar{A}/\sigma$, where $\bar{A}$ is the sample mean of $A$.

### 5.2.2 Performance metrics

Evaluating the quality of denoising in medical imaging is of subjective nature, as it is often based on the particular requirements of a medical expert. Nevertheless, there are a number of standard evaluation metrics used in the literature, some of which we adopt in the present

study. Specifically, one such metric is the *root mean square error* (RMSE), which can be expressed in dB as

$$\mathrm{RMSE} = 20\log_{10}\left[\frac{1}{N}\sum_{k=1}^{N}|e_k|^2\right]^{1/2}, \tag{5.3}$$

where $e_k = A_k - \hat{A}_k$ denotes the difference between the original intensity $A_k$ and its estimated value $\hat{A}_k$ at sample $k$, and $N$ stands for the total number of image samples.

When estimate $\hat{A}_k$ is biased, the mean value of $e_k$ may not be equal to zero, in general. In this case, it makes sense to replace $e_k$ in (5.3) by its centred version $e_k - \bar{e}$, with $\bar{e}$ being the sample mean of $e_k$ given by

$$\bar{e} = \frac{1}{N}\sum_{k=1}^{N} e_k. \tag{5.4}$$

The resulting metric is called the *centred* RMSE (cRMSE), and it is formally defined as

$$\mathrm{cRMSE} = 20\log_{10}\left[\frac{1}{N}\sum_{k=1}^{N}|e_k - \bar{e}|^2\right]^{1/2}. \tag{5.5}$$

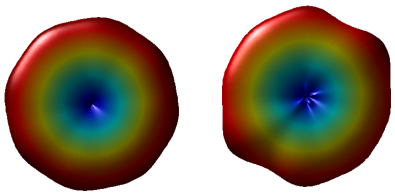
It should be noted that, while structurally similar, the RMSE and cRMSE metrics provide different quantitative assessment in the case of biased estimation. Consequently, the analysis and comparison of both these metrics can be helpful in evaluating the performance of the de-biasing procedures detailed in Section 4.4.

### 5.2.3 Comparative analysis of algorithm performance

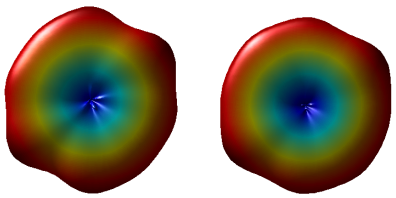
#### Weight adaptivity and fibre direction estimation of proposed methods

The adaptive property of NLM weights has been a prime motivation for developing NLM methods for HARDI. In this section we briefly compare the performance of the two proposed NLM filters with that of the linear GKS filtering approach using synthetic HARDI data. Spherical functions, corresponding to the HARDI signal at a fixed location, were generated with differing relative smoothness. In particular, two sets of signals were fabricated, both with a $S_0 = 60$, b-value of 3000 s/mm$^2$ and $K = 64$ samples. The signals in the first set consisted of a single "fibre" with FA = 0.9, while the signals in the second set consisted of three crossing fibres having fractional anisotropy (FA) of 0.7. Examples of the above signals are shown in Subplots 5.1(a) and 5.2(a), respectively (note that displayed signals

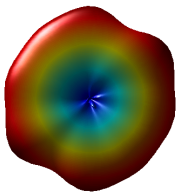
have been interpolated using spherical harmonics to provide an image resolution greater than 64 samples per function). All the test signals were contaminated by Rician noise to give rise to an SNR ($\bar{A}/\sigma$) of 10. The noisy versions of the test signals are shown in the Subplots 5.1(b) and 5.2(b).



(a) Original (b) Noisy

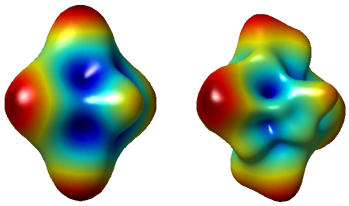


(c) GKS (d) RILES

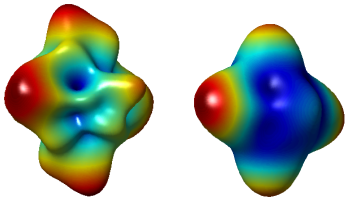


(e) NLMS2

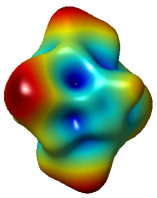
Figure 5.1: HARDI signal from set 1 (FA = 0.9)



(a) Original (b) Noisy



(c) GKS (d) RILES



(e) NLMS2

Figure 5.2: HARDI signal from set 2 (FA = 0.7)

Subplots (c)-(e) in Figures 5.1 and 5.2 show the denoised estimates of the GKS, RILES, and NLMS2 methods, respectively. Results have been compared in terms of RMSE, computed based on averaging the results of 100 independent trials, and are summarized in Table 5.2 with the best results recorded in bold. Side experiments have shown that when GKS parameters were tuned to filter one type of dataset, the method performed competitively. However, as both the figures and Table 5.2 show, when the parameters were set to produce results beneficial to both types of HARDI signals, the performance was

Table 5.2: Results of adaptivity experiment

| | Set 1 RMSE (dB) | Set 2 RMSE (dB) |
|---|---|---|
| Noisy | 4.57 | -1.29 |
| GKS | 4.17 | -1.50 |
| RILES | **2.16** | **−3.85** |
| NLMS2 | 4.51 | -2.12 |

suboptimal.

Unlike GKS, the NLMS2 and RILES weights are not fixed, but are instead adaptively chosen based on neighbourhood similarity. As a result, there is no guarantee as to the number of samples which will be given a significant weight in the average. We can, however, predict that because a) RILES is based on spherical harmonic band energies and b) numerous neighbourhood structures can have the same band energy, that c) RILES will be more liberal than NLMS2 in distributing large weights. This characteristic is beneficial in the current experiment where there are only 64 samples, and even fewer samples with the same original intensity. As can be seen from the figures, the small number of similar samples has resulted in NLMS2 performing a minimal amount of averaging. RILES, on the other hand, still performed a significant amount of denoising. Although the strict similarity measure of NLMS2 is better able to filter out dissimilar (erroneous) samples from the average, the image must contain a greater number of similar samples for this method to be effective.

The first set of signals (Figure 5.1) was also used to evaluate the proposed methods' performance in improving the accuracy of fibre direction estimation. The true direction of each synthetic fibre was compared to the main eigenvector of the DTI diffusion tensor for noisy and denoised signals. The results are compared in terms of angular error in degrees, computed based on averaging the results from 100 independent trials, and are summarized in Table 5.3 with the best results recorded in bold. It is noted that both proposed methods are able to increase the accuracy of fibre direction estimation.

Table 5.3: Results of fibre direction estimation

| | Set 1 error of direction estimation (degrees) |
|---|---|
| Noisy | 1.90 |
| GKS | 1.83 |
| RILES | **0.76** |
| NLMS2 | 1.26 |

**Algorithm performance on 16x16 phantom**

The experiments reported in this section use 16x16 phantom whose noise-free reconstruction is shown in Subplot (a) of Figures 5.3, 5.5, and 5.7. The phantom is a 2D grid of spherical HARDI functions emulating two crossing fibres superimposed on one circular fibre. All spherical functions were constructed using the multi-tensor model with $S_0 = 60$, a b-value of 3000 s/mm$^2$, and $K = 64$ samples. The phantom was subjected to three levels of Rician noise giving rise to SNRs of 5, 10, and 20 shown in Subplot (b) of Figures 5.3, 5.5, and 5.7, respectively. For the sake of visual comparison, Subplots (c)-(f) of the same figures show the denoised estimates of the vTV, NLMv, RILES, and NLMS2 algorithms respectively, whereas Figures 5.4,5.4, and 5.4 show the corresponding OPDFs implemented using the method of [89].

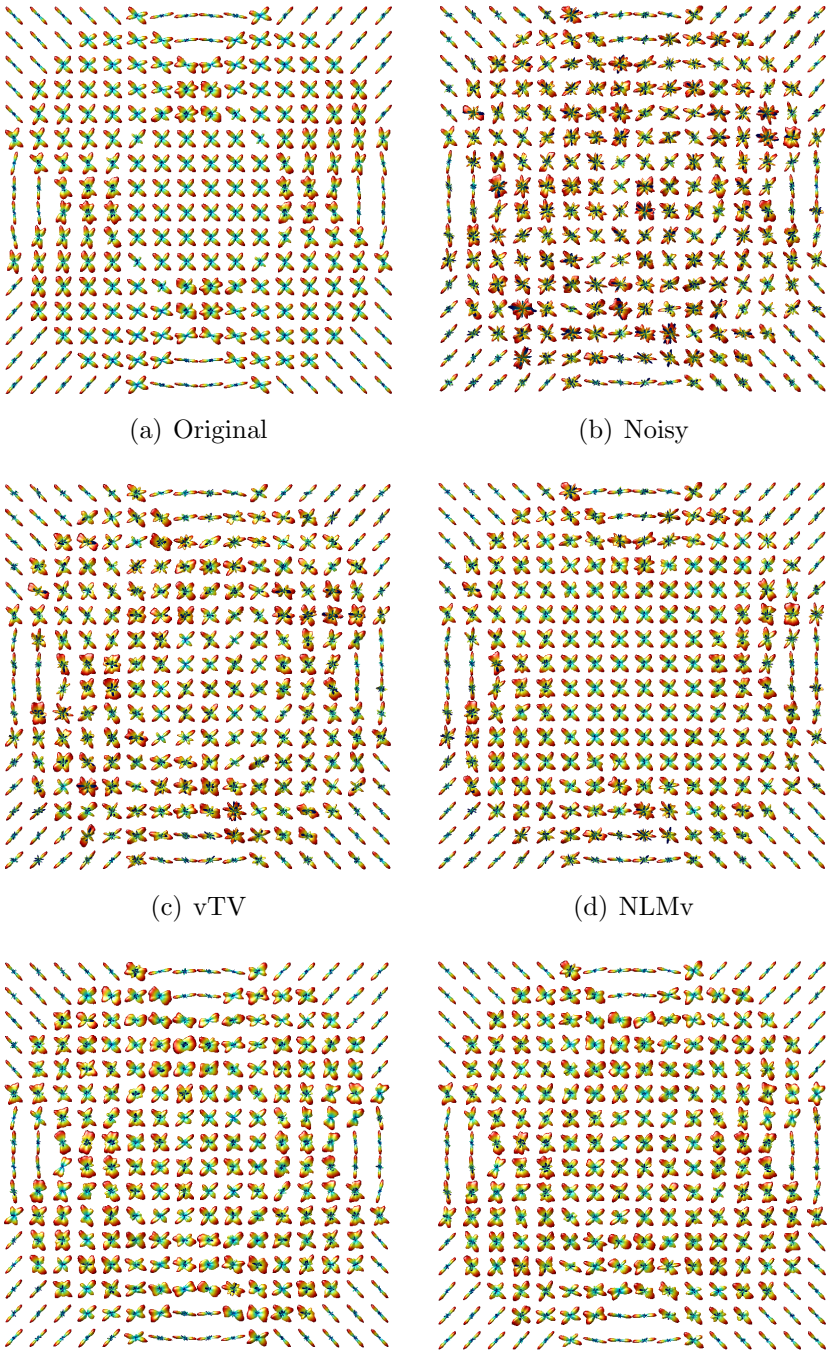
Table 5.4 summarizes the values of the performance metrics obtained by evaluating each algorithm's estimate of raw HARDI data and reconstructed OPDFs for differing SNRs (best results are reported in bold). It can be seen that, among all the compared methods, NLMS2 provides better performance in terms of all the performance measures. An additional important observation can be made through comparing the values of the RMSE and cRMSE metrics obtained using different reconstruction algorithms. In particular, we first note that the values of RMSE and cRMSE corresponding to the noisy data are not identical – a fact which indicates the presence of a non-zero bias in the measurement noise. Whereas the NLMv, RILES, and NLMS2 denoising techniques have approximately equal RMSE and cRMSE values, the measures are noticeably different in the case of vTV. This fact suggests that the latter method is inefficient in removing the constant bias in the reconstruction error.

(a) Original (b) Noisy

(c) vTV (d) NLMv

(e) RILES (f) NLMS2

Figure 5.3: HARDI signal reconstruction for SNR 5

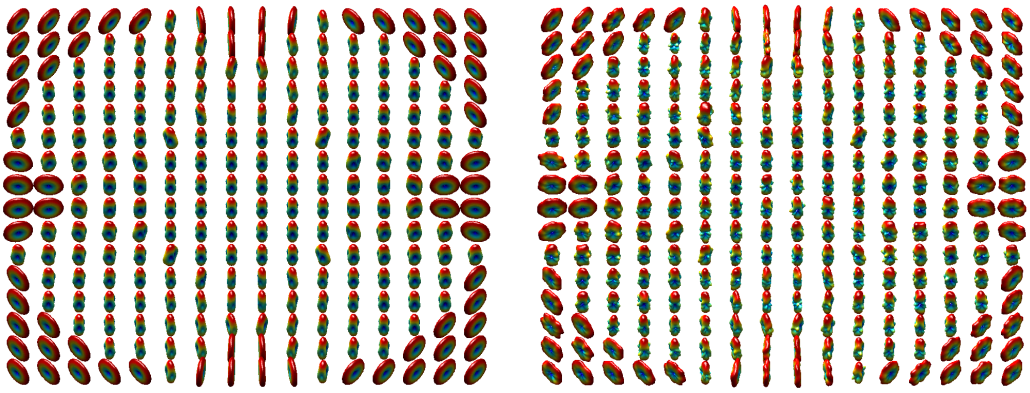


(a) Original (b) Noisy

(c) vTV (d) NLMv

(e) RILES (f) NLMS2

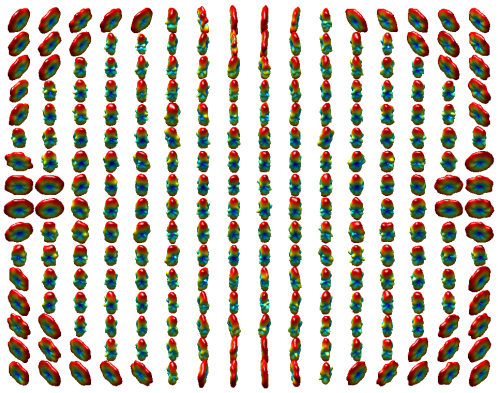
Figure 5.4: OPDF reconstruction for SNR 5



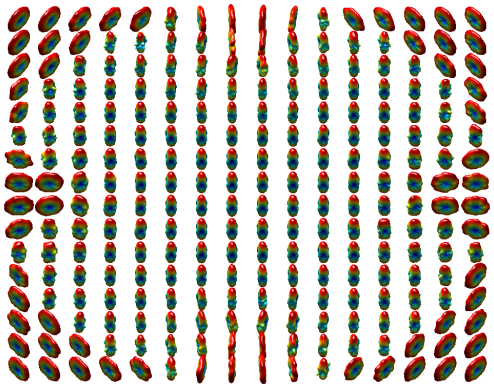
(a) Original



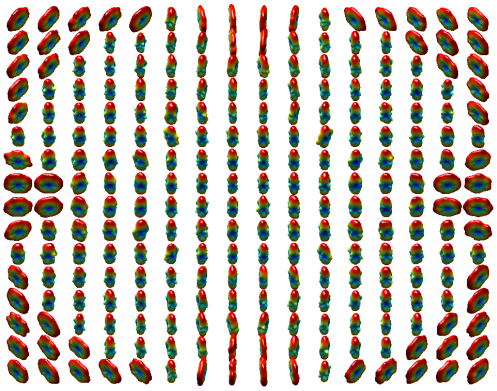
(b) Noisy



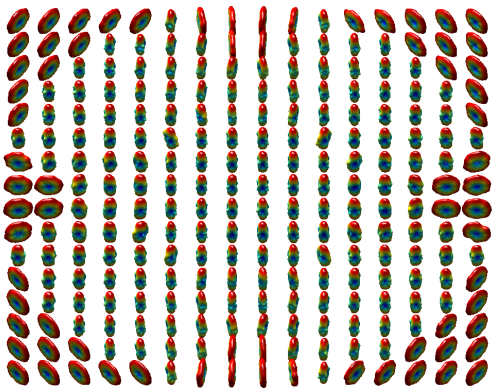
(c) vTV



(d) NLMv

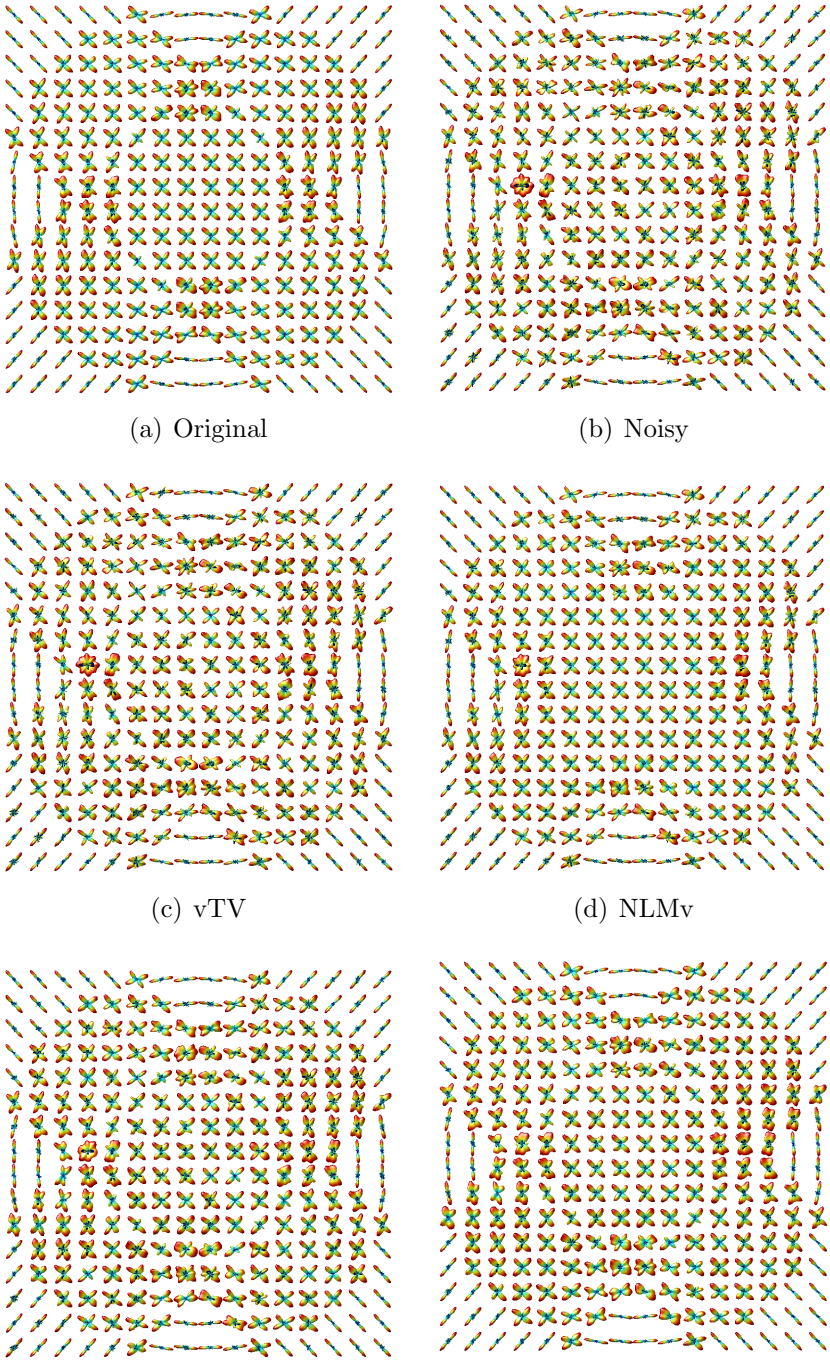


(e) RILES



(f) NLMS2

Figure 5.5: HARDI signal reconstruction for SNR 10

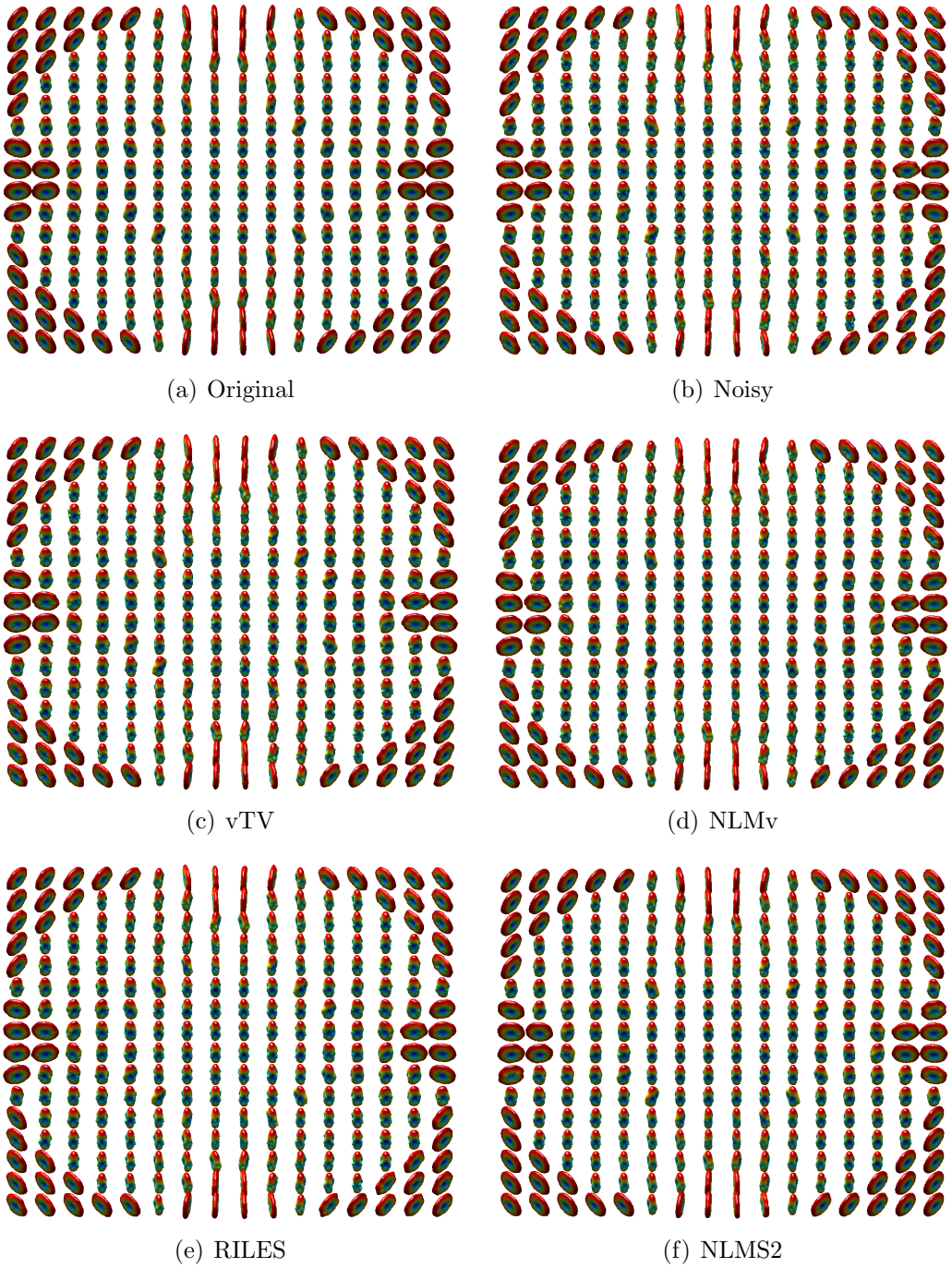


(a) Original (b) Noisy

(c) vTV (d) NLMv

(e) RILES (f) NLMS2

Figure 5.6: OPDF reconstruction for SNR 10

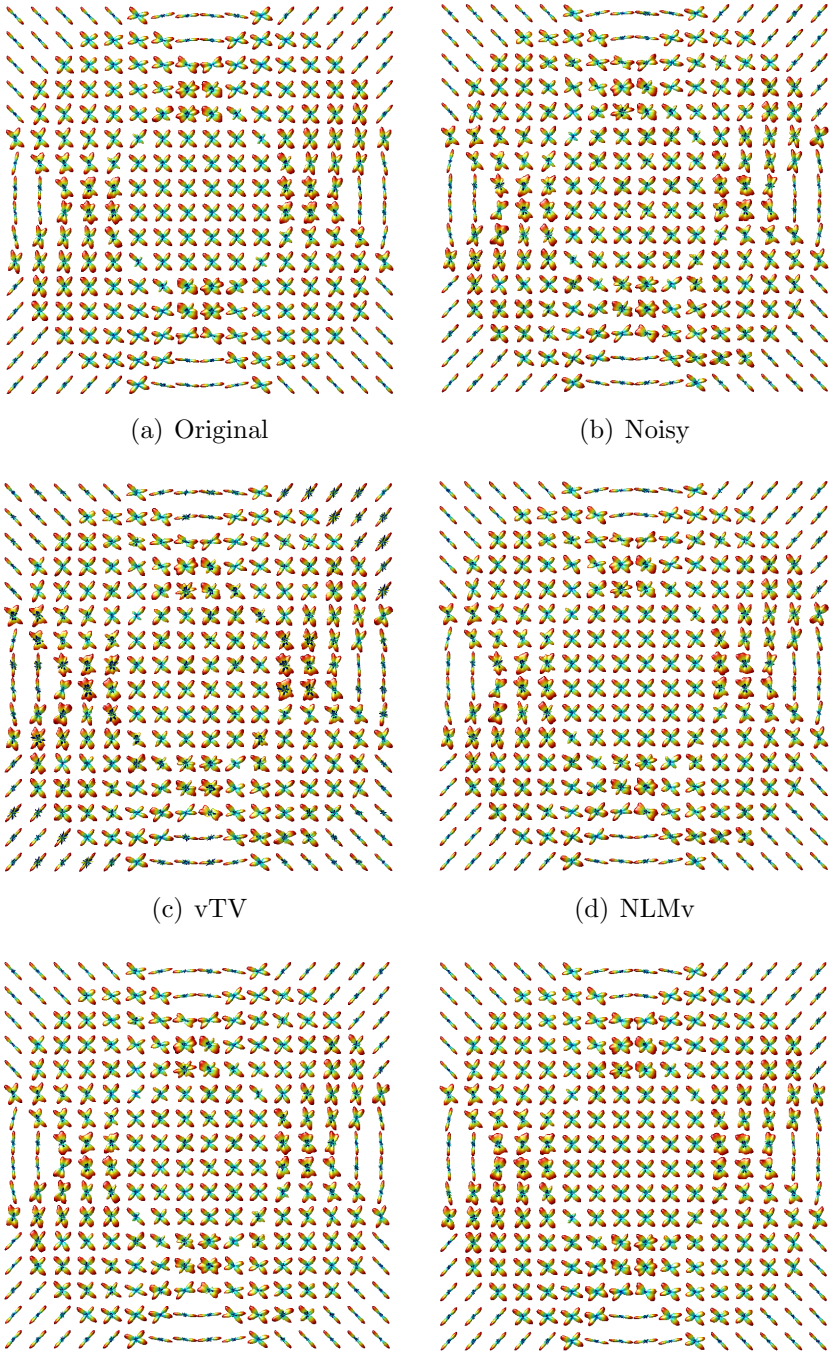


(a) Original (b) Noisy

(c) vTV (d) NLMv

(e) RILES (f) NLMS2

Figure 5.7: HARDI signal reconstruction for SNR 20



(a) Original (b) Noisy

(c) vTV (d) NLMv

(e) RILES (f) NLMS2

Figure 5.8: OPDF reconstruction for SNR 20

Table 5.4: Results of denoising 16x16 synthetic phantom

| | | Raw HARDI Data | | OPDF Reconstruction |
|---|---|---|---|---|
| | SNR ($\bar{A}/\sigma$) | RMSE (dB) | cRMSE (dB) | RMSE (dB) |
| Noisy | 5 | 14.3503 | 14.2036 | -4.7855 |
| vTV | | 11.9712 | 11.9679 | -5.2005 |
| NLMv | | 10.7916 | 10.7817 | -5.2770 |
| RILES | | 10.2617 | 10.2616 | -5.5861 |
| NLMS2 | | **9.6967** | **9.6965** | **−5.6020** |
| Noisy | 10 | 8.2780 | 8.2067 | -5.1660 |
| vTV | | 8.0923 | 8.0755 | -5.1966 |
| NLMv | | 6.9742 | 6.8934 | -5.3158 |
| RILES | | 5.7489 | 5.7475 | -5.2875 |
| NLMS2 | | **4.3851** | **4.3748** | **−5.4060** |
| Noisy | 20 | 2.2678 | 2.2551 | -5.3203 |
| vTV | | 1.9174 | 1.9167 | -4.4515 |
| NLMv | | 1.6363 | 1.6332 | -5.3330 |
| RILES | | 0.9354 | 0.9310 | -5.3285 |
| NLMS2 | | **−1.3443** | **−1.3443** | **−5.3713** |

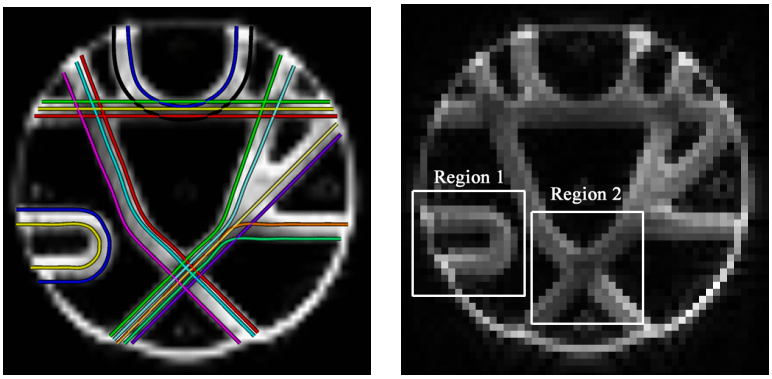
## 5.3 Fibre cup phantom

The Fibre Cup contest was held by the Medical Image Computing and Computer-Assisted Intervention (MICCAI) Society in 2009 to evaluate reconstruction models and tractography algorithms in diffusion MRI. An MR phantom was constructed containing a plethora of realistic crossing, kissing, splitting and bending fibre configurations to be used as a ground truth dataset for method comparison (the fibre locations can be seen in Figure 5.9(a)). The data set was later made available to the public for the purpose of assessing new processing methods. We will use the phantom to visually assess the proposed RILES and NLMS2 filtering methods.

Diffusion-weighted data of the phantom was acquired on the 3T Tim Trio MRI systems of the NeuroSpin centre. A single-shot diffusion-weighted twice refocused spin echo echo-planar pulse sequence was used to perform the acquisitions. The voxel size was 3x3x3 $mm^3$ and the b-value was 2000 $s/mm^2$. Diffusion sensitization was applied along a set of 64 orientations, uniformly distributed over the sphere.

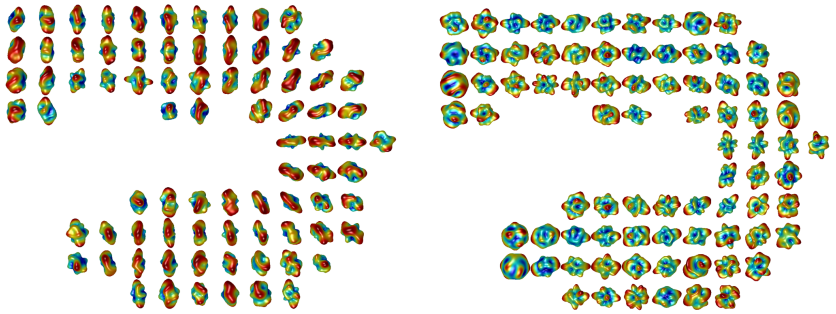
The denoising results are displayed in two magnified regions of the phantom, the locations of which are shown in Figure 5.9(b). Figure 5.10 shows the results for the first

region, which contains a U fibre, whereas the results for the crossing fibres in the second region are shown in Figure 5.11. A visual assessment verifies that both methods preserve anatomical structure, while allowing for a clearer view of diffusion directions.




(a) Fibre cup ground truth (b) Locations of displayed results

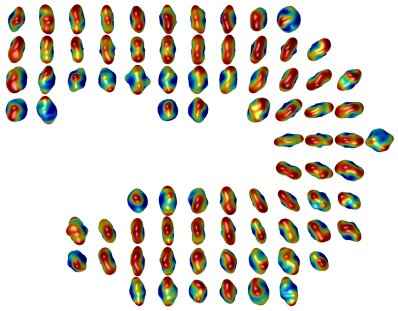
Figure 5.9: Fibre cup phantom



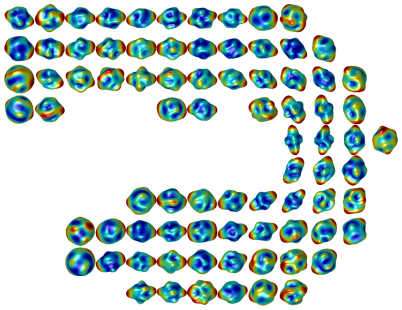
(a) Noisy Data



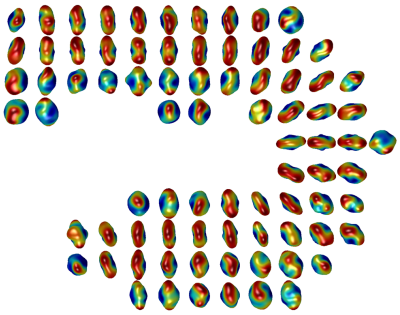
(b) Noisy OPDF



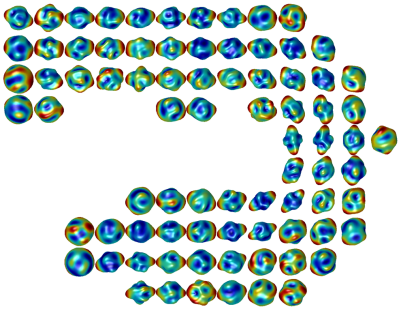
(c) RILES Data



(d) RILES OPDF

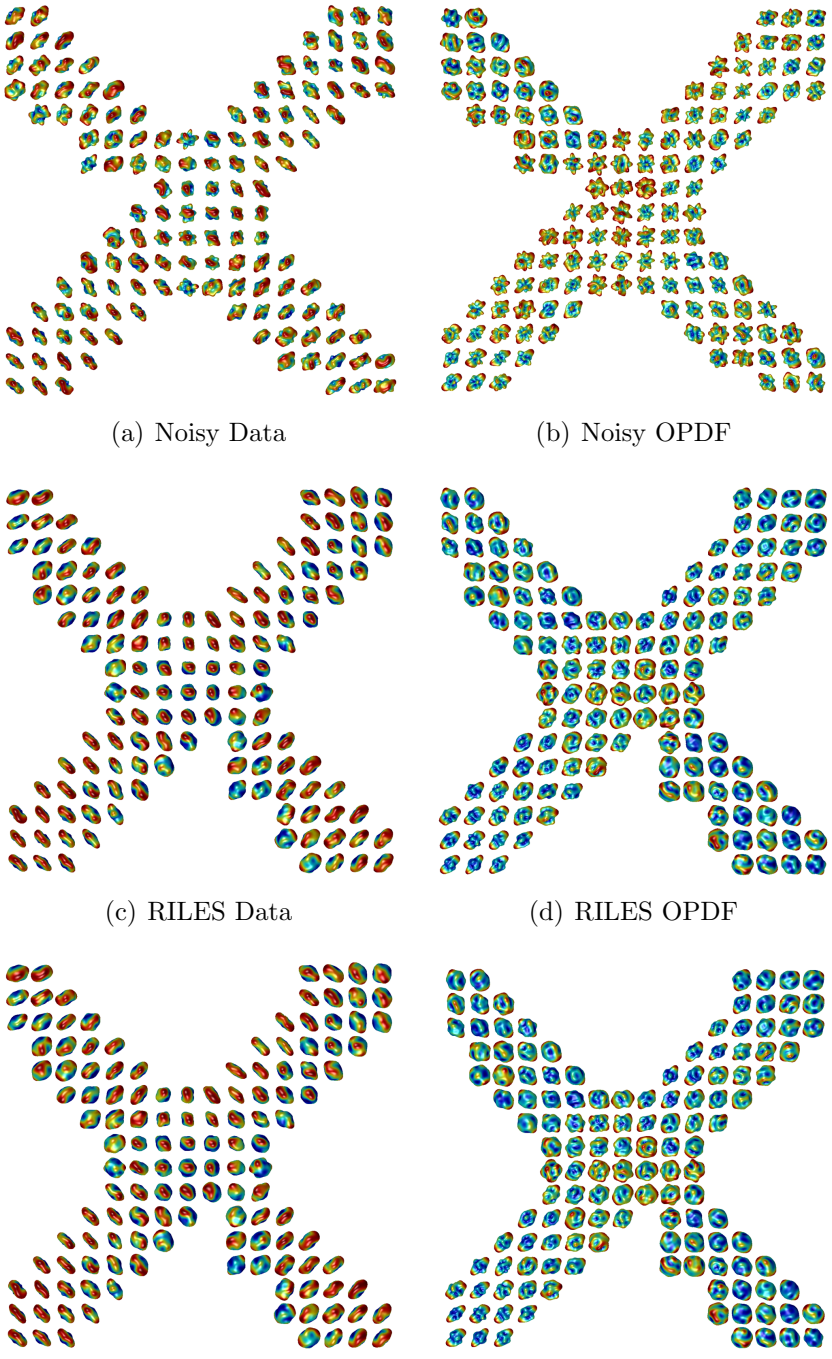


(e) NLMS2 Data



(f) NLMS2 OPDF

Figure 5.10: Fibre cup denoised region 1



(a) Noisy Data

(b) Noisy OPDF

(c) RILES Data

(d) RILES OPDF

(e) NLMS2 Data

(f) NLMS2 OPDF

Figure 5.11: Fibre cup denoised region 2

## 5.4 Experiments with real-life data

In this section, we visually assess the ability of the proposed RILES and NLMS2 filtering techniques to denoise diffusion-weighted measurements of a human brain. The data was kindly provided by Robarts Research Institute, acquired from a Siemens 3T scanner using a single-shot diffusion-weighted once refocused spin echo echoplanar pulse sequence. The voxel size was 2x2x2 $mm^3$ and the b-value was 1000 $s/mm^2$. Diffusion sensitization was applied along a set of 64 orientations.

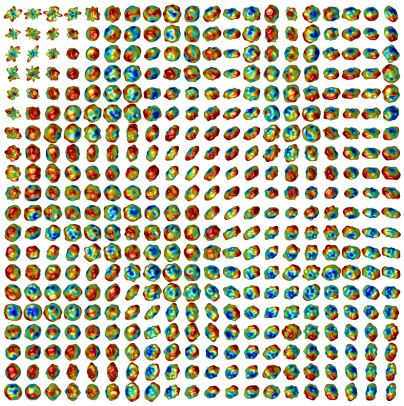
Figure 5.12 displays the ability of diffusion-weighted MRI techniques to capture the morphology of the white matter invisible to traditional contrasts. Additionally, it is clear from this figure the impact that the proposed denoising methods (in this case RILES) have on the OPDF reconstruction. Magnified denoising results are displayed in three regions of the brain, the locations of which are shown in Figure 5.12(a). Figures 5.13, 5.14, and 5.15 show the denoising results of the three regions, respectively. A visual assessment verifies that both methods preserve anatomical structure, while increasing the ability to delineate neural fibres.


(a) $B_0$ image (locations of displayed results)

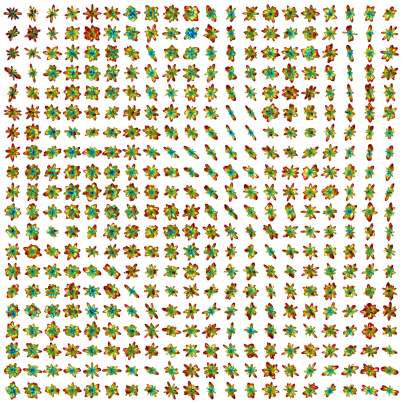
(b) Diffusion-weighted MRI (noisy OPDF)

(c) Diffusion-weighted MRI (denoised OPDF)

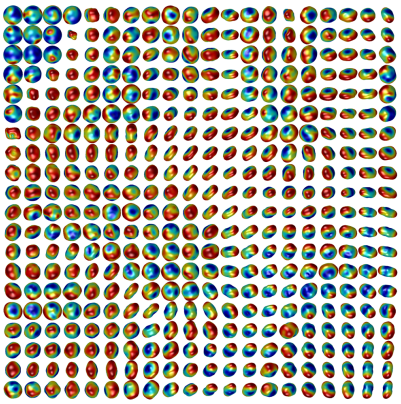
Figure 5.12: Diffusion-weighted MRI of the human brain



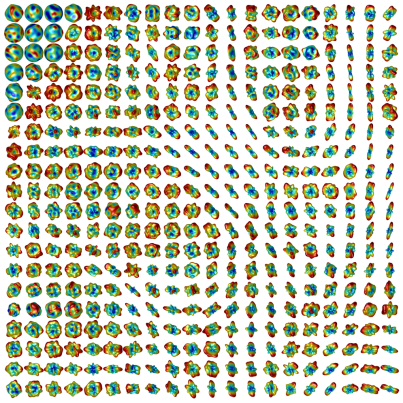
(a) Noisy Data



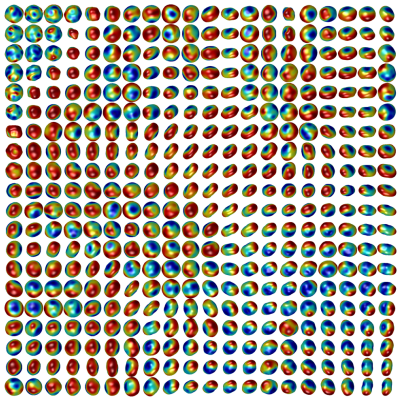
(b) Noisy OPDF



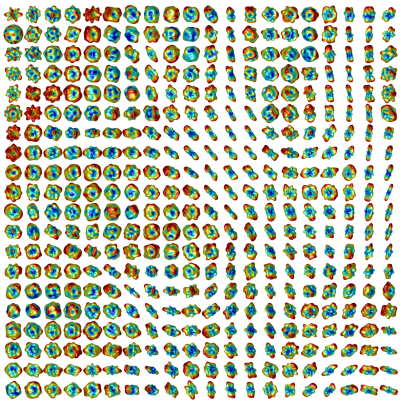
(c) RILES Data



(d) RILES OPDF

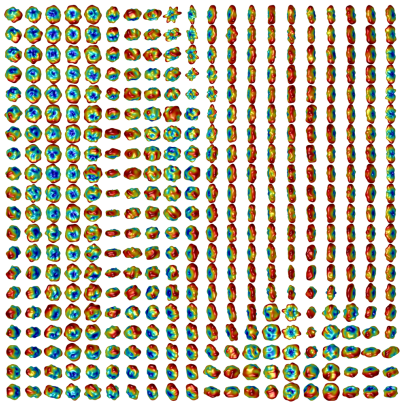


(e) NLMS2 Data

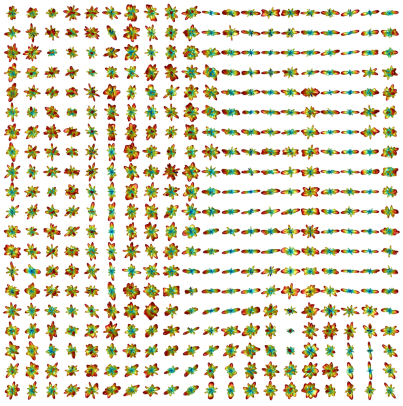


(f) NLMS2 OPDF

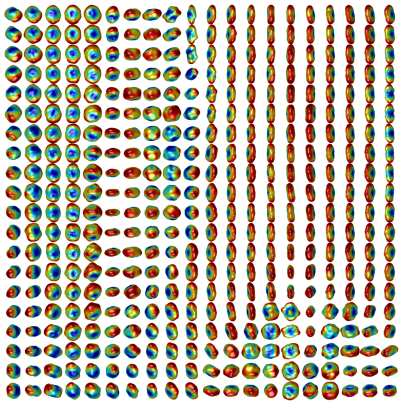
Figure 5.13: Brain denoised region 1



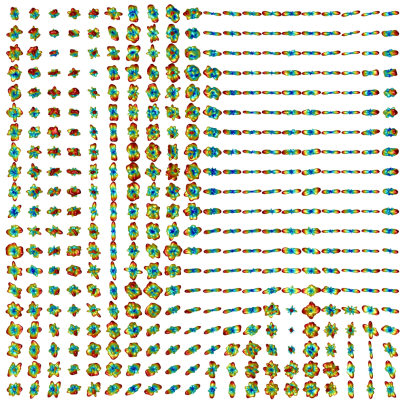
(a) Noisy Data



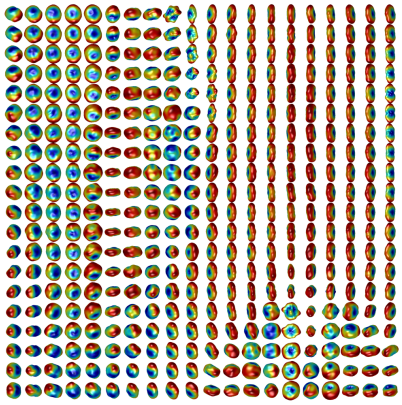
(b) Noisy OPDF



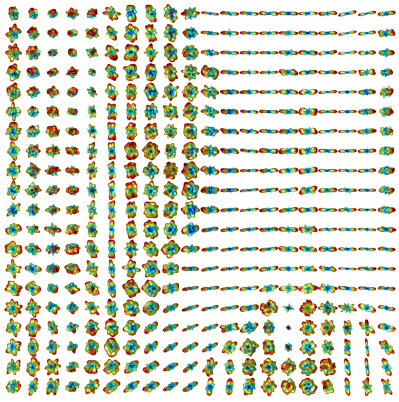
(c) RILES Data



(d) RILES OPDF

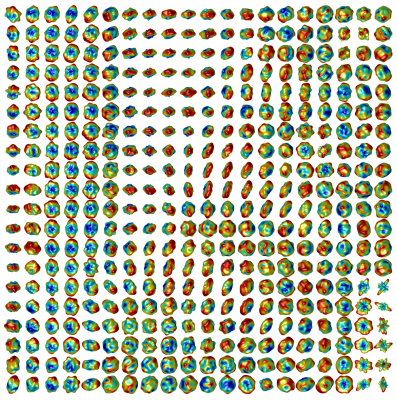


(e) NLMS2 Data

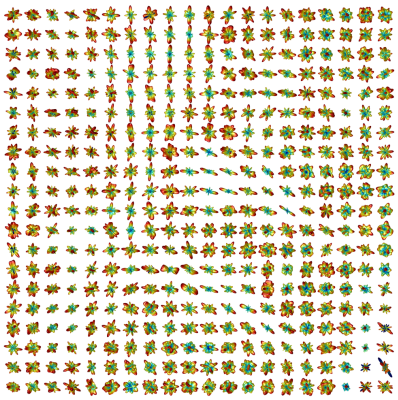


(f) NLMS2 OPDF

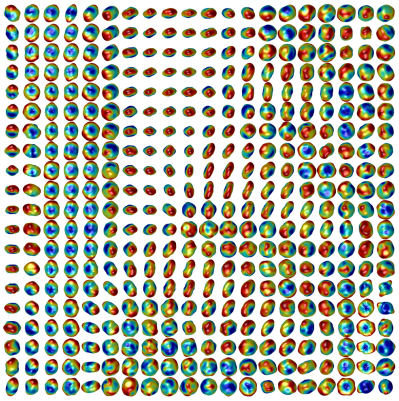
Figure 5.14: Brain denoised region 2



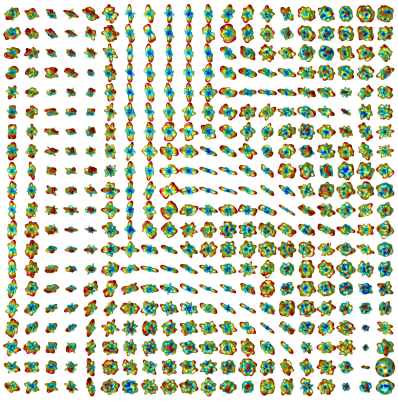
(a) Noisy Data



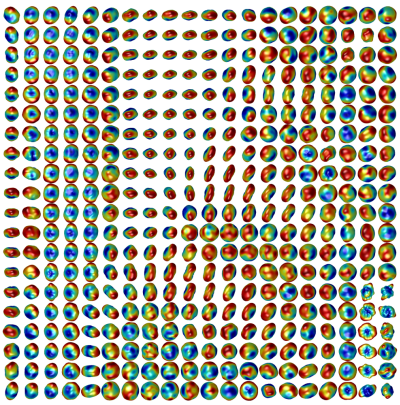
(b) Noisy OPDF



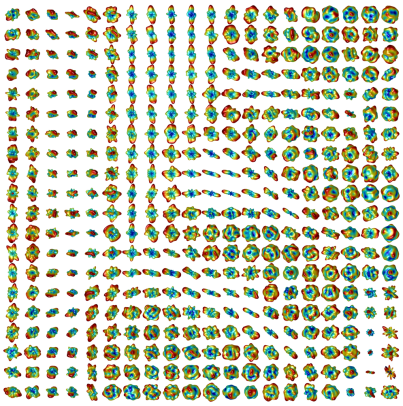
(c) RILES Data



(d) RILES OPDF



(e) NLMS2 Data



(f) NLMS2 OPDF

Figure 5.15: Brain denoised region 3

# Chapter 6

# Conclusions and Future Work

The aim of this work is to develop an effective framework for the filtering of HARDI measurement noise in order to correct the distortion of diagnostically relevant imaging details. The framework should take into account both the manifold to which the HARDI signal belongs and the statistical nature of MRI measurement noise. These goals were accomplished using an approach rooted in non-local means (NLM) weighted averaging. The average included samples, and therefore dependencies, from the entire manifold and the result of the average, being an approximation of the noisy measurement's expected value, was used to deduce an estimate of the original signal value in accordance with MRI statistics.

Unlike linear filters, the NLM averaging weights are determined adaptively based on a neighbourhood similarity measure. This approach allows a more accurate estimate of the noisy data's expected value. The particular method of neighbourhood comparison proposed in this thesis is one of spherical neighbourhoods, which assigns large weights to samples with similar local orientational diffusion characteristics. Given the complicated nature of spherical neighbourhood comparison, two methods have been proposed. The NLMS2 algorithm introduces a sample-to-neighbourhood similarity measure as a precursor of neighbourhood-to-neighbourhood comparison. This method, while having a rigorous theoretical background, is computationally intensive. The RILES algorithm, on the other hand, compares neighbourhoods using spherical harmonic band energies. Although less rigorous, RILES is more computationally efficient than NLMS2.

Both of the proposed algorithms have been evaluated on simulated and real data. Results from simulated experiments have shown that the proposed methods have better performance in comparison to other reference methods developed for HARDI denoising.

The NLMS2 algorithm consistently outperformed all other methods in denoising HARDI images with various signal-to-noise ratios, while RILES provided a computationally efficient alternative with competitive results. Visual results obtained from experiments in which the proposed methods were applied to real data have shown an increased ability to delineate the anatomical structure of fibres from the corresponding ODFs.

Despite the success of the proposed methods, there may still be the opportunity for improvement. In particular, the chosen weighting kernel $\beta_k$ in NLMS2 and, analogously, the Gauss-Weierstrass windowing kernel chosen in RILES may not be optimal. Further experiments should be conducted comparing other weighting techniques. Moreover, alternative methods of $\mathbb{S}^2$ neighbourhood comparison should be explored. For instance, another method for neighbourhood comparison can be derived from the zero phase spherical harmonic coefficients. Since the zero phase coefficients correspond to zonal basis functions, they provide a neighbourhood description that is invariant to azimuth rotations. Alternatively, one could compare neighbourhoods using all the spherical harmonic coefficients. Although this method is no longer invariant to azimuth rotations, it provides a neighbourhood comparison that is more strict than RILES and may be able to improve performance by more accurately determining which samples are truly similar. Finally, further research should be conducted concerning methods which add spatial dimensions into the neighbourhood, as was outlined in Section 4.3.2.

In conclusion, the proposed NLM methodology has been shown to efficiently filter HARDI noise. Although further improvements may be possible, the methods presented in this thesis can currently be used as a preprocessing step to improve the results of the numerous processing techniques available to HARDI.

# Appendix A

# Derivation of Fourier relationship between $E(\mathbf{q})$ and $\overline{P}(\mathbf{r})$

This appendix provides the derivation (following reference [25]) of the Fourier relationship between Ensemble Average Propagator and PGSE diffusion signal. For simplicity, k-space considerations are initially ignored and only included in a later section.

## A.1 PGSE

In the PGSE experiment, the diffusion-encoding gradients are pulses (Figure A.1). The pulses are assumed to be sufficiently narrow to neglect any spin displacement over their duration, allowing a clear distinction between the encoding time $\delta$ and the diffusion time $\Delta$.

Consider a collection of spins in volume V undergoing a PGSE experiment. Under the narrow pulse approximation, a spin at location $\mathbf{r}$ subjected to the first gradient pulse accumulates a phase

$$\phi_{\mathbf{r}} = \gamma\delta\mathbf{g}\cdot\mathbf{r}. \tag{A.1}$$

If the spin then moves to location $\mathbf{r}'$ during $\Delta$, application of the second gradient pulse reverses the phase according to

$$\phi_{\mathbf{r}'} = -\gamma\delta\mathbf{g}\cdot\mathbf{r}'. \tag{A.2}$$

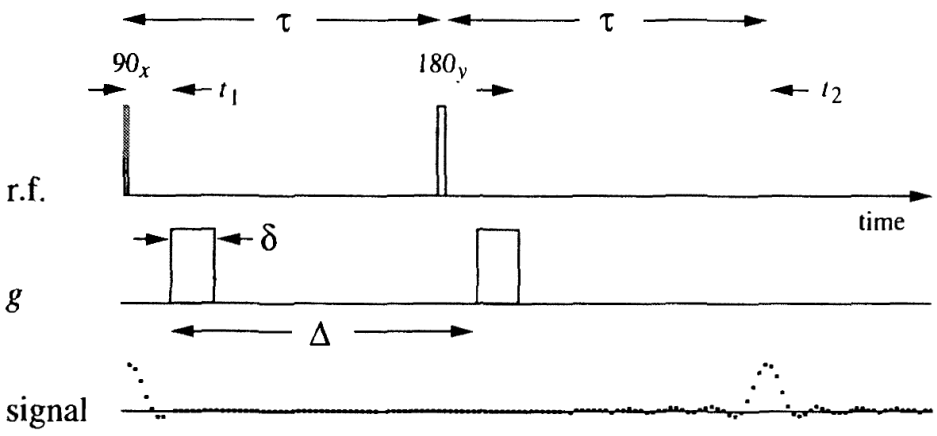



Figure A.1: PGSE sequence [25]

The net phase change that a spin undergoes is

$$\phi_{\mathbf{r}'} + \phi_{\mathbf{r}} = -\gamma\delta\mathbf{g}\cdot(\mathbf{r}' - \mathbf{r}). \tag{A.3}$$

For stationary spins, the net phase change is zero and the signal is unaffected. If the spin moves, then the second gradient does not correctly compensate for $\phi_{\mathbf{r}}$. The result is spin dephasing causing signal attenuation.

The demodulated signal can be written as (see Equation 3.13) and Equation 3.15)

$$s(t) \propto \int |\mathbf{M}_{\perp}(\mathbf{r},t)| e^{-i\phi'(\mathbf{r},t)} d\mathbf{r}, \quad \text{where } \phi'(\mathbf{r},t) = -\gamma\delta\mathbf{g}\cdot(\mathbf{r}'(t) - \mathbf{r}), \tag{A.4}$$

where $\mathbf{g}$ is a diffusion-encoding gradient causing the dephasing captured in $\phi'(\mathbf{r},t)$. In this context, it is convenient to denote the signal $s(t)$ as $S(\mathbf{g})$ to make explicit the dependence on the encoding gradient.

In the present case, only the diffusion attenuation effect is of interest. Therefore, it is convenient to consider the signal independent of the spin density, T1 relaxation, and T2 relaxation effects contained in $|\mathbf{M}_{\perp}(\mathbf{r},t)|$. To eliminate these effects, the diffusion attenuated signal $S(\mathbf{g})$ is normalized by a signal $S(\mathbf{0})$ obtained in the absence of encoding gradients and diffusion attenuation

$$E(\mathbf{g}) = S(\mathbf{g})/S(\mathbf{0}). \tag{A.5}$$

## A.2 Q-space and the Average Propagator

The signal amplitude of $E(\mathbf{g})$ is governed solely by the net phase change $\phi'(\mathbf{r}, t)$. Yet, the net phase change is a function of random spin movements. Consequently, $E(\mathbf{g})$ should be rewritten in probabilistic terms.

To this end, let $P_s(\mathbf{r}|\mathbf{r}', t)$ represent the chance that a single spin placed at location $\mathbf{r}$ diffuses to location $\mathbf{r}'$ after a time $t$. The average dephasing at a given location $\mathbf{r}'$ is

$$\int P_s(\mathbf{r}|\mathbf{r}', t) e^{i\gamma\delta\mathbf{g}\cdot(\mathbf{r}'-\mathbf{r})} d\mathbf{r}. \tag{A.6}$$

Equation A.6 assumes that all starting locations are equally likely; there is no consideration that the likelihood of a spin arriving from a starting location $\mathbf{r}$ depends on the initial spin density at $\mathbf{r}$. Taking this into account, let $\rho(\mathbf{r})$ be the initial spin density normalized by an appropriate factor such that

$$\int\int \rho(\mathbf{r}) P_s(\mathbf{r}|\mathbf{r}', t) d\mathbf{r} d\mathbf{r}' = 1. \tag{A.7}$$

$P_s(\mathbf{r}|\mathbf{r}', t)$ accounts for the likelihood that a "diffusion force" takes a single spin from $\mathbf{r}$ to $\mathbf{r}'$ whereas $\rho(\mathbf{r})$ accounts for the likelihood that a spin starts at $\mathbf{r}$. The received signal can be written as an ensemble average in which each phase term $e^{i\gamma\delta\mathbf{g}\cdot(\mathbf{r}'-\mathbf{r})}$ is weighted by the probability for any spin in the volume to begin at $\mathbf{r}$ and diffuse to $\mathbf{r}'$

$$E_\Delta(\mathbf{g}) = \int\int \rho(\mathbf{r}) P_s(\mathbf{r}|\mathbf{r}', \Delta) e^{i\gamma\delta\mathbf{g}\cdot(\mathbf{r}'-\mathbf{r})} d\mathbf{r} d\mathbf{r}'. \tag{A.8}$$

Introducing the variable $\mathbf{q} = (2\pi)^{-1}\gamma\delta\mathbf{g}$, $E_\Delta(\mathbf{g})$ can be rewritten in Q-space as

$$E_\Delta(\mathbf{q}) = \int\int \rho(\mathbf{r}) P_s(\mathbf{r}|\mathbf{r}', \Delta) e^{i2\pi\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})} d\mathbf{r} d\mathbf{r}'. \tag{A.9}$$

At this point it is convenient to introduce the Ensemble Average Propagator (EAP). The Ensemble Average Propagator gives the average probability for a spin in volume V to diffuse with a net displacement $\mathbf{R} = \mathbf{r}' - \mathbf{r}$ over a time t. This probability is given by

$$\overline{P_s}(\mathbf{R}, t) = \int \rho(\mathbf{r}) P_s(\mathbf{r}|\mathbf{r} + \mathbf{R}, t) d\mathbf{r}. \tag{A.10}$$

Using the EAP, it is not difficult to see that through a change of variables Equation A.9 can be rewritten as

$$E_\Delta(\mathbf{q}) = \int \overline{P_s}(\mathbf{R}, \Delta) e^{i2\pi \mathbf{q}\cdot\mathbf{R}} d\mathbf{R}. \tag{A.11}$$

It is evident that recovery of the EAP is performed via the Fourier Transform. Q-space can be traversed by varying the pulse duration $\delta$ or encoding gradient strength. Thus, by taking enough samples in Q-space we can recover a diffusion profile for the spins in volume V. Rather than obtain the diffusion profile of the entire volume V, it is desirable to obtain the diffusion profile at each voxel. This is accomplished by combining the Q-space with K space developments.

## A.3 Q-space and K space

Recovering the EAP at each voxel requires the spatial resolution obtained from phase encoding. A k-space imaging gradient is applied in addition to the PGSE gradients resulting in the signal

$$S(\mathbf{k}, \mathbf{q}) = \int \rho(\mathbf{r}) e^{i2\pi \mathbf{k}\cdot\mathbf{r}} \int P_s(\mathbf{r}|\mathbf{r}', \Delta) e^{i2\pi \mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})} d\mathbf{r}' d\mathbf{r}. \tag{A.12}$$

Which can be written

$$S(\mathbf{k}, \mathbf{q}) = \int \rho(\mathbf{r}) E_\Delta(\mathbf{q}, \mathbf{r}) e^{i2\pi \mathbf{k}\cdot\mathbf{r}} d\mathbf{r}, \tag{A.13}$$

where

$$E_\Delta(\mathbf{q}, \mathbf{r}) = \int P_s(\mathbf{r}|\mathbf{r}', \Delta) e^{i2\pi \mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})} d\mathbf{r}'. \tag{A.14}$$

It is clear that reconstruction in k-space recovers an image with voxel intensity $\rho(\mathbf{r})E_\Delta(\mathbf{q}, \mathbf{r})$. Each image reconstructed from k-space provides all voxels with another measurement in **q**-space. If we acquire enough images with different **q** values, every voxel to use the **q**-Fourier Transform to recover $\rho(\mathbf{r})P_s(\mathbf{r}|\mathbf{r}', \Delta)$ - the EAP for an infinitesimal volume surrounding **r**. However, in practice the volume surrounding **r** is not infinitesimal, it is a voxel of finite size. Therefore, it is more accurate to consider the **q**-Fourier Transform of $\rho(\mathbf{r})E_\Delta(\mathbf{q}, \mathbf{r})$ as the local EAP averaged over the volume of a voxel. In summary, by acquiring enough DW images to properly cover q-space, it is possible to recover the EAP at each voxel.

# Appendix B

# Derivation of statistical sample similarity measures for MRI

## B.1 Subtractive sample similarity measure for non-central chi square statistics

We are interested in evaluating $p_{F_{s-k}-F_{t-k}|G_{s-k},G_{t-k}}$ at 0, which can be rewritten as [85]

$$p_{F_{s-k}-F_{t-k}|G_{s-k},G_{t-k}}(0 \mid g_{s-k}, g_{t-k}) = \int_0^\infty p_{F_{s-k}|G_{s-k},G_{t-k}}(f \mid g_{s-k}, g_{t-k}) p_{F_{t-k}|G_{s-k},G_{t-k}}(f \mid g_{s-k}, g_{t-k})\, df \tag{B.1}$$

where it has been assumed that $F_{s-k}$ and $F_{t-k}$ are conditionally independent, given the noisy data $G_{s-k}$ and $G_{t-k}$. Moreover, further simplification using (4.14) and an application of Bayes theorem leads to,

$$p_{F_{s-k}-F_{t-k}|G_{s-k},G_{t-k}}(0 \mid g_{s-k}, g_{t-k}) = \frac{\int_0^\infty p_{F_{s-k}}(f) p_{F_{t-k}}(f) p_{G_{s-k}|F_{s-k}}(g_{s-k} \mid f) p_{G_{t-k}|F_{t-k}}(g_{t-k} \mid f)\, df}{p_{G_{s-k}}(g_{s-k}) p_{G_{t-k}}(g_{t-k})} \tag{B.2}$$

If no prior information about the original intensity is provided, one can assume a uniform density on the original intensities [30, 85]. Therefore, one can move $p_{F_{s-k}}(f)$ and $p_{F_{t-k}}(f)$

out of the integral. Also, $p_{G_{s-k}}(g_{s-k})$ can be written as

$$\begin{aligned} p_{G_{s-k}}(g_{s-k}) &= \int_0^\infty p_{G_{s-k}|F_{s-k}}(g_{s-k} \mid f_{s-k}) p_{F_{s-k}}(f)\, df \\ &= p_{F_{s-k}}(f) \int_0^\infty \frac{1}{2} e^{-(g_{s-k}+f)/2} I_0(\sqrt{f g_{s-k}})\, df \\ &= p_{F_{s-k}}(f) \end{aligned} \tag{B.3}$$

Moving from the first line to the second uses the uniform density assumption of $F_{s-k}$. Moving from the second line to the third uses the fact that integrating conditional probability $p_{G_{s-k}|F_{s-k}}(g_{s-k} \mid f_{s-k})$ with respect to $f$ is equivalent to integrating with respect to $g_{s-k}$, which evaluates to one. In the same manner, it can be shown that

$$p_{G_{t-k}}(g_{t-k}) = p_{F_{t-k}}(f) \tag{B.4}$$

Using (B.3) and (B.4), (B.2) simplifies to

$$\begin{aligned} p_{F_{s-k}-F_{t-k}|G_{s-k},G_{t-k}}(0 \mid g_{s-k}, g_{t-k}) &= \int_0^\infty p_{G_{s-k}|F_{s-k}}(g_{s-k} \mid f) p_{G_{t-k}|F_{t-k}}(g_{t-k} \mid f)\, df \\ &= \frac{1}{4} e^{-(g_{s-k}+g_{t-k})/2} \int_0^\infty e^{-f} I_0(\sqrt{g_{s-k} f}) I_0(\sqrt{g_{t-k} f})\, df \end{aligned} \tag{B.5}$$

Performing a change of variables to $y = \sqrt{f}$, one obtains

$$\begin{aligned} p_{F_{s-k}-F_{t-k}|G_{s-k},G_{t-k}}(0|g_{s-k}, g_{t-k}) &= \frac{1}{2} e^{-(g_{s-k}+g_{t-k})/2} \int_{y=0}^\infty y e^{-y^2} I_0(\sqrt{g_{s-k}} y) I_0(\sqrt{g_{t-k}} y)\, dy \\ &= \frac{1}{2} e^{-(g_{s-k}+g_{t-k})/2} \int_{y=0}^\infty y e^{-y^2} J_0(i\sqrt{g_{s-k}} y) J_0(i\sqrt{g_{t-k}} y)\, dy \end{aligned} \tag{B.6}$$

where $J_0$ is the Bessel function of the first kind and order 0, and $i = \sqrt{-1}$. Subsequently, using equation (2.32) of [42],

$$\begin{aligned} p_{F_{s-k}-F_{t-k}|G_{s-k},G_{t-k}}(0|g_{s-k}, g_{t-k}) &= \frac{1}{4} e^{-(g_{s-k}+g_{t-k})/4} I_0\left(-\frac{\sqrt{g_{s-k} g_{t-k}}}{2}\right) \\ &= \frac{1}{4} e^{-(g_{s-k}+g_{t-k})/4} I_0\left(\frac{\sqrt{g_{s-k} g_{t-k}}}{2}\right) \end{aligned} \tag{B.7}$$

where the last line makes use of $I_0(-x) = I_0(x)$.

## B.2 Rational sample similarity measure for Rician statistics

In this case, we are interested in evaluating $p_{A_{s-k}/A_{t-k}|M_{s-k},M_{t-k}}$ at 1, which can be rewritten as [85]

$$p_{A_{s-k}/A_{t-k}|M_{s-k},M_{t-k}}(1 \mid m_{s-k}, m_{t-k}) = \int_0^\infty a p_{A_{s-k}|M_{s-k}}(a \mid m_{s-k}) p_{A_{t-k}|M_{t-k}}(a \mid m_{t-k})\, da \tag{B.8}$$

Proceeding similar to the derivation in Appendix B.1, we obtain,

$$p_{A_{s-k}/A_{t-k}|M_{s-k},M_{t-k}}(1 \mid m_{s-k}, m_{t-k}) = \int_0^\infty a p_{M_{s-k}|A_{s-k}}(m_{s-k} \mid a) p_{M_{t-k}|A_{t-k}}(m_{t-k} \mid a)\, da =$$
$$= \frac{m_{s-k} m_{t-k}}{\sigma^4} e^{-(m_{s-k}^2 + m_{t-k}^2)/2\sigma^2} \int_{a=0}^\infty a e^{-a^2/\sigma^2} I_0\left(\frac{m_{s-k} a}{\sigma^2}\right) I_0\left(\frac{m_{t-k} a}{\sigma^2}\right)\, da \tag{B.9}$$

Performing a change of variables to $y = a/\sigma$, one obtains

$$p_{A_{s-k}/A_{t-k}|M_{s-k},M_{t-k}}(1 \mid m_{s-k}, m_{t-k}) = \tag{B.10}$$
$$\frac{m_{s-k} m_{t-k}}{\sigma^2} e^{-(m_{s-k}^2 + m_{t-k}^2)/2\sigma^2} \int_0^\infty y e^{-y^2} I_0\left(\frac{m_{s-k}}{\sigma} y\right) I_0\left(\frac{m_{t-k}}{\sigma} y\right)\, dy$$

and using formula (2.32) of [42], this evaluates to

$$p_{A_{s-k}/A_{t-k}|M_{s-k},M_{t-k}}(1 \mid m_{s-k}, m_{t-k}) = \frac{m_{s-k} m_{t-k}}{\sigma^2} e^{-(m_{s-k}^2 + m_{t-k}^2)/4\sigma^2} I_0\left(\frac{m_{s-k} m_{t-k}}{\sigma^2}\right) \tag{B.11}$$

# Bibliography


[1] I. Aganj, C. Lenglet, and G. Sapiro. Odf reconstruction in q-ball imaging with solid angle consideration. In *Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging: From Nano to Macro*, ISBI'09, pages 1398–1401. IEEE Press, 2009. 32

[2] A. L. Alexander, J. E. Lee, M. Lazar, and A. S. Field. Diffusion tensor imaging of the brain. *Neurotherapeutics*, 4:316–329, 2007. 3, 26

[3] D. C. Alexander. Multiple-fiber reconstruction algorithms for diffusion MRI. *Annals of the New York Academy of Science*, 1064:113–133, 2005. 4, 27

[4] D. C. Alexander, G. J. Barker, , and S. R. Arridge. Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Journal of Magnetic Resonance Imaging*, 48(2):331–340, 2002. 4, 27, 30, 34

[5] C. S. Anand and J. S. Sahambi. Wavelet domain non-linear filtering for MRI denoising. *Magnetic Resonance Imaging*, 28(6):842–861, 2010. 8

[6] A. W. Anderson. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. *Magnetic Resonance in Medicine*, 54:1194–1206, 2005. 4, 28, 30

[7] K. Arfanakis, B. Hermann, B. Rogers, J. Carew, M. Seidenberg, and M. Meyerand. Diffusion tensor MRI in temporal lobe epilepsy. *Magnetic Resonance Imaging*, 20(7):511–519, 2002. 1, 3, 24

[8] P. Bao and L. Zhang. Noise reduction for magnetic resonance images via adaptive multiscale products thresholding. *IEEE Transactions on Medical Imaging*, 22(9):1089–1099, September 2003. 8

[9] P. J. Basser, J. Mattiello, and D. LeBihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. Imag., Ser. B*, 103(3):247–254, 1994. 1, 3, 26, 27

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

[11] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44:625–632, 2000. 3, 26

[12] S. Basu, T. Fletcher, and R. Whitaker. Rician noise removal in Diffusion Tensor MRI. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 117–125, 2006. 8

[13] T. Behrens, M. Woolrich, M. Jenkinson, H. Johansen-Berg, R. Nunes, S. Clare, P. Matthews, J. Brady, and S. Smith. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, 50:1077–1088, 2003. 3, 26

[14] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. MR imaging of intravoxel incoherent motions: Application to diffusion and perfusion in neurological disorders. *Radiology*, 161:401–407, 1986. 1, 3, 26

[15] D. Le Bihan, J.-F. Mangin, C. Poupon, C. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13:534–546, 2001. 3, 26

[16] D. Le Bihan, C. Poupon, A. Amadon, and F. Lethimonnier. Artifacts and pitfalls in diffusion MRI. *Journal of Magnetic Resonance Imaging*, 24:478–488, 2006. 4

[17] P. Blomgren and T. F. Chan. Color tv: Total variation methods for restoration of vector valued images. *IEEE Trans. Image Processing*, 7:304–309, 1996. 11, 12

[18] J. Boulanger, C. Kevrann, and P. Bouthemy. Space-time adaptation for patch-based image sequence restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1096–1102, June 2007. 8

[19] D. Brunet, E. R. Vrscay, and Z. Wang. Structural similarity-based affine approximation and self-similarity of images revisited. In *Proceedings of the 8th international conference on Image analysis and recognition - Volume Part II*, ICIAR'11, pages 264–275, Berlin, Heidelberg, 2011. Springer-Verlag. 38

[20] A. Buades, B. Coll, and J. M. Morel. A non local algorithm for image denoising. In *IEEE International Conference on Computer Vision and Pattern Recognition, CVPR 2005*, volume 2, pages 60–65, 2005. 8, 13, 38, 39

[21] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, 4(2):490–530, 2005. 8, 13

[22] A. Buades, B. Coll, and J. M. Morel. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76(2):123–139, 2008. 8, 13

[23] A. Buades, B. Coll, and J. M. Morel. Image denoising methods. a new nonlocal principle. *SIAM Rev.*, 52:113–147, February 2010. 38

[24] T. Bülow. Spherical diffusion for 3D surface smoothing. In *Proceedings of 1st International Symposium on Data Processing Visualization and Transmission (3DPVT2002)*, pages 449–458, June 2002. 9, 54

[25] P. T. Callahan. *Principles of nuclear magnetic resonance microscopy*. Clarendon Press, Oxford, UK, 1991. 3, 14, 17, 20, 28, 79, 80

[26] M. K. Chung, S. M. Robbins, K. M. Dalton, R. J. Davidson, A. L. Alexander, and A. C. Evans. Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, 25(4):1256 – 1265, 2005. 9, 54, 58

[27] G.G. Cleveland, D.C. Chang, C.F. Hazlewood, and H.E. Rorschach. Nuclear magnetic resonance measurement of skeletal muscle: anisotrophy of the diffusion coefficient of the intracellular water. *Biophysical Journal*, 16(9):1043 – 1053, 1976. 26

[28] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Transactions on Medical Imaging*, 27(4):425–441, March 2008. 8

[29] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, August 2007. 8

[30] C. A. Deledalle, L. Denis, and F. Tupin. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. *IEEE Transactions on Image Processing*, 18(12):2661–2672, December 2009. 39, 40, 44, 45, 46, 83

[31] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. Apparent diffusion coefficients from high angular resolution diffusion images: Estimation and applications. *Magnetic Resonance in Medicine*, 56(2):395–410, 2006. 4, 28, 30

[32] 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. 30, 34

[33] M. Descoteaux, R. Deriche, D. Le Bihan, J.-F. Mangin, and C. Poupon. Diffusion propagator imaging: Using Laplace's equation and multiple shell acquisitions to reconstruct the diffusion propagator. In *Lecture Notes in Computer Science*, volume 5636, pages 1–13. Springer, 2009. 3

[34] M. Descoteaux, N. Wiest-Daesslé, S. Prima, C. Barillot, and R. Deriche. Impact of rician adapted non-local means filtering on hardi. In *Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention, Part II*, MICCAI '08, pages 122–130. Springer-Verlag, 2008. 8

[35] G. Dhatt and G. Touzot. *The finite element method displayed*. (A Wiley-Interscience publication). Wiley, 1984. 11

[36] C. M. Ellis, A. Simmons, D. K. Jones, J. Bland, J. M. Dawson, M. A. Horsfield, S. C. Williams, and P. N. Leigh. Diffusion tensor MRI assesses corticospinal tract damage in ALS. *Neurology*, 53:1051–1058, 1999. 1, 3, 24

[37] A. Fan, W. M. Wells, J.W. Fisher, M. Cetin, S. Haker, R. Mulkem, C. Tempany, and A. S. Willsky. A unified variational approach to denoising and bias correction in MR. *Inf. Process Med Imag*, 18:148–159, 2003. 8

[38] M. Filippi, M. Cercignani, M. Inglese, M. A. Horsfield, and G. Comi. Diffusion tensor magnetic resonance imaging in multiple sclerosis. *Neurology*, 56:304–311, 2001. 1, 3, 24

[39] L. Frank. Anisotropy in high angular resolution diffusion-tensor MRI. *Magnetic Resonance in Medicine*, 45:935–939, 2001. 4, 27

[40] L. R. Frank. Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic Resonance in Medicine*, 47:1083–1099, 2002. 4, 27, 28, 30, 34

[41] G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, June 1992. 8

[42] M. Lawrence Glasser and Emilio Montaldi. Some integrals involving Bessel functions. *Journal of Mathematical Analysis and Applications*, 183(1):577–590, May 1994. 84, 85

[43] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, third edition, 2001. 40

[44] H. Groemer. *Geometric applications of Fourier series and spherical harmonics*. Cambridge University Press, 1996. 34

[45] H. Gudbjartsson and S Patz. The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6):910–914, December 1995. 33

[46] M. E. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley-Liss, 1999. 14, 17, 19, 20, 21

[47] J. R. Hansen. Pulsed nmr study of water mobility in muscle and brain tissue. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 230(3):482 – 486, 1971. 26

[48] D. Healy and J. Weaver. Two applications of wavelet transforms in magnetic resonance imaging. *IEEE Transactions on Information Theory*, 38(2):840–860, March 1992. 8

[49] M. Hilton, T. Ogden, D. Hattery, G. Eden, and B. Jawerth. Wavelet denoising of functional MRI data. In A. Aldroubi and M. Unser, editors, *Wavelets in Medicine and Biology*, pages 93–114. CRC press, 1996. 8

[50] K. M. Jansons and D. C. Alexander. Persistent angular structure: New insights from diffusion magnetic resonance imaging data. *Inverse Problems*, 19:1031–1046, 2003. 30

[51] T M Jessell and E R Kandel. Synaptic transmission: a bidirectional and self-modifiable form of cell-cell communication. *Cell*, 72 Suppl:1–30, Jan 1993. 1

[52] H. Johansen-Berg and T. E. J. Behrens. *Diffusion MRI: From quantitative measurements to in-vivo neuroanatomy*. Academic Press, first edition, 2009. 1, 2, 3, 6, 17

[53] D. K. Jones. The effect of gradient sampling schemes on measures derived from Diffusion Tensor MRI: A Monte Carlo study. *Magnetic Resonance in Medicine*, 51:807–815, 2004. 3, 26

[54] D. K. Jones and P. J. Basser. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 52(5):979–993, November 2004. 4, 6

[55] C. Kervrann, J. Boulanger, and P. Coupé. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In *In Proc. Conf. Scale-Space and Variational Meth. (SSVM 07*, 2007. 40

[56] Y. Kim, P. M. Thompson, A. W. Toga, L. Vese, and L. Zhan. Hardi denoising: Variational regularization of the spherical apparent diffusion coefficient sadc. In *Proceedings of the 21st International Conference on Information Processing in Medical Imaging*, IPMI '09, pages 515–527. Springer-Verlag, 2009. 12, 58

[57] M. Lysaker, A. Lundervold, and X. C. Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590, December 2003. 8

[58] A. Macovski. Noise in MRI. *Magnetic Resonance in Medicine*, 36(3):494–497, 1996. 33

[59] J. Malcolm, M. Shenton, and Y. Rathi. Filtered tractography: State estimation in a constrained subspace. In *Proceedings of MICCAI (Diffusion Modeling and Fiber Cup)*, pages 122–133, 2009. 3, 4, 26

[60] J. Malcolm, M. Shenton, and Y. Rathi. Neural tractography using an unscented Kalman filter. In *Proceedings of IPMI*, pages 126–138, 2009. 3, 4, 26

[61] J. G. Malcolm, O. Michailovich, S. Bouix, C.-F. Westin, A. Tannenbaum, M. E. Shenton, and Y. Rathi. An unscented Kalman filtering approach to neural tractography using the Watson directional function. *Medical Image Analysis*, 14(1):58–69, 2009. 3, 4, 26

[62] J. V. Manjóm, P. Coupé, L. Marti-bonmati, M. Robles, and D. L. Collins. Adaptive non-local means denoising of MR images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging*, 31:192–203, 2010. 8

[63] J. V. Manjon, J. Carbonell-Caballero, J. J. Lull, G. G. Martí, L. Martí-Bonmatí, and M. Robles. MRI denoising using non-local means. *Medical Image Analysis*, 12:514–523, 2008. 8

[64] T. McGraw, B. Vemuri, E. Ozarslan, Y. Chen, and T. Mareci. Variational denoising of diffusion weighted MRI. *Inverse Probl. Imaging*, 3(4):625–648, 2009. 10, 11

[65] O. Michailovich and Y. Rathi. Approximation of orientation distribution functions by spherical ridgelets. In *Proceedings of ISBI*, Paris, 2008. 30

[66] S. Mori. *Introduction to diffusion tensor imaging*. Elsevier, 2007. 3, 26

[67] S. Mori, K. Frederiksen, P. C. M. Van Zijl, B. Stieltjes, M. A. Kraut, M. Solaiyappan, and M. G. Pomper. Brain white matter anatomy of tumor patients evaluated with diffusion tensor imaging. *Annals of Neurology*, 51(3):377–380, 2002. 1, 2, 3

[68] T. Moritani, S. Ekholm, and P.-L.Westesson. *Diffusion-weighted MR imaging of the brain*. Springer-Verlag, 2005. 1

[69] O. Naggara, C. Oppenheim, D. Rieu, N. Raoux, S. Rodrigo, G. Dalla Barba, and J. Meder. Diffusion tensor imaging in early Alzheimer's disease. *Psychiatry Research: Neuroimaging*, 146(3):243–249, 2006. 1, 3, 24

[70] R.D. Nowak. Wavelet-based Rician noise removal for Magnetic Resonance Imaging. *IEEE Transactions on Image Processing*, 8(10):1408–1419, October 1999. 8

[71] E. Ozarslan, B. C. Vemuri, and T. H. Mareci. Generalized scalar measures for diffusion MRI using trace, variance, and entropy. *Magnetic resonance in medicine*, 53(4):866–876, April 2005. 11

[72] P. Perona and J. Malik. Scale-spave and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990. 8

[73] C. Pierpaoli, P. Jezzard, P. Basser, A. Barnett, and G. Di Chiro. Diffusion tensor MR imaging of the human brain. *Radiology*, 201(3):637–648, 1996. 3, 26

[74] A. Pizurica, W. Philips, L. Lemahieu, and M. Acheroy. A versatile wavelet domain noise filtration technique for medical imaging. *IEEE Transactions on Medical Imaging*, 22(3):323–331, March 2003. 8

[75] A. Pizurica, A. M. Wink, E. Vansteenkiste, W. Philips, and J. B. T. M. Roerdink. A review of wavelet denoising in MRI and ultrasound brain imaging. *Current Medical Imaging Reviews*, 2(2):247–260, May 2006. 8

[76] Y. Rathi, J. G. Malcolm, O. Michailovich, R. W. McCarley, C.-F. Westin, and M. Shenton. Biomarkers for identifying first-episode schizophrenia patients using diffusion weighted imaging. In *Lecture Notes in Computer Science*, volume 6361, Beijing, China, September 2010. MICCAI. 3

[77] P. Rigoard, K. Buffenoir, N. Jaafari, J. P. Giot, J. L. Houeto, P. Mertens, S. Velut, and B. Bataille. The accumbofrontal fasciculus in the human brain: a microsurgical anatomical study. *Neurosurgery*, 68(4):1102–1111; discussion 1111, 2011. 48

[78] S. E. Rose, A. L. Janke, and J. B. Chalk. Gray and white matter changes in Alzheimer's disease: A diffusion tensor imaging study. *Journal of Magnetic Resonance Imaging*, 27(1):20–26, 2007. 1, 3, 24

[79] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4):259–268, 1992. 11

[80] J. Sijbers, A. J. Dekker, P. Scheunders, and D. V. Dyck. Maximum-likelihood estimation of Rician distribution parameters. *IEEE Transactions on Medical Imaging*, 17(3):357–361, June 1998. 56

[81] S. M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T. E. Nichols, C. E. Mackay, K. E. Watkins, O. Ciccarelli, M. Z. Cader, P. M. Matthews, and T. E. Behrens. Tractbased spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31:1487–1505, 2006. 3, 26

[82] E. O. Stejskal. Use of Spin Echoes in a Pulsed Magnetic-Field Gradient to Study Anisotropic, Restricted Diffusion and Flow. *The Journal of Chemical Physics*, 43:3597–3603, 1965. 26

[83] E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: Spin echoes in the presence of time-dependent field gradient. *Journal of Chemical Physics*, 42(1):288–292, January 1965. 2, 3, 24, 26

[84] P. C. Sundgren, Q. Dong, D. Gomez-Hassan, S. K. Mukherji, P. Maly, and R. Welsh. Diffusion tensor imaging of the brain: Review of clinical applications. *Neuroradiology*, 46(5):339–350, 2004. 1, 3, 24

[85] T. Teuber and A. Lang. A new similarity measure for nonlocal filtering in the presence of multiplicative noise. *Preprint University of Kaiserslautern*, 2011. 39, 41, 43, 44, 45, 83, 85

[86] P. Tofts. *Quantitative MRI of the brain: Measuring changes caused by disease*. John Wiley and Sons, 2003. 1

[87] H. C. Torrey. Bloch Equations with Diffusion Terms. *Physical Review*, 104:563–565, 1956. 25

[88] J.-D. Tournier, F. Calamante, D. G. Gadian, and A. Connelly. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23:1176–1185, 2004. 30

[89] A. Tristan-Vega, C. F. Westin, and S. Aja-Fernandez. Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging. *NeuroImage*, 47:638–650, 2009. 32, 62

[90] D. S. Tuch. *Diffusion MRI of complex tissue structure*. PhD thesis, Harvard, 2002. 28

[91] D. S. Tuch. Q-ball imaging. *Magnetic Resonance in Medicine*, 52:1358–1372, 2004. 30, 31

[92] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine*, 48:577–582, 2002. 4, 27, 28

[93] D. S. Tuch, T. G. Reese, M. R. Wiegell, and V. J. Wedeen. Diffusion MRI of complex neural architecture. *Neuron*, 40:885–895, 2003. 4, 27, 28, 30, 31

[94] J. B. Weaver, Yansun Xu, D. M. Healy Jr., and L. D. Cromwell. Filtering noise from images with wavelet transforms. *Magnetic Resonance in Medicine*, 21(2):288–295, November 1991. 8

[95] V. J. Wedeen, P. Hagmann, W.-Y. Tseng, T. G. Reese, and R. M. Weisskoff. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine*, 54:1377–1386, 2005. 3, 28

[96] C. Westbrook, C. K. Roth, and J. Talbot. *MRI in practice*. Wiley-Blackwel, 2005. 1

[97] C.-F. Westin, S. Maier, H. Mamata, A. Nabavi, F. Jolesz, and R. Kikinis. Processing and visualization of diffusion tensor MRI. *Medical Image Analysis*, 6(2):93–108, 2002. 3

[98] M. R. Wiegell, H. Larsson, and Van J. Wedeen. Fiber crossing in human brain depicted with diffusion tensor MR imaging. *Radiology*, 217:897–903, 2000. 3, 4, 26, 27

[99] N. Wiest-Daesslé, S. Prima, P. Coupé, S. P. 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 (MICCAI)*, pages 344–351, 2007. 8, 13, 48, 55, 58

[100] N. Wiest-Daesslé, S. Prima, P. Coupé, S. P. Morrissey, and C. Barillot. Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI. In *Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention, Part II*, MICCAI '08, pages 171–179, Berlin, Heidelberg, 2008. Springer-Verlag. 8, 56

[101] J. C. Wood and K. M. Johnson. Wavelet packet denoising of magnetic resonance images: Importance of Rician noise at low SNR. *Magnetic Resonance in Medicine*, 41(3):631–635, March 1999. 8

[102] T. A. Woolsey, J. Hanaway, and M. H. Gado. *The brain atlas: A visual guide to the human nervous system.* John Wiley and Sons, Hoboken, NJ, third edition, 2008. 2

[103] Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu. Wavelet transform domain filters: A spatially selective noise filtration technique. *IEEE Transactions on Image Processing*, 3(6):747–758, November 1994. 8

[104] A. Yershova and S. M. Lavalle. Deterministic Sampling Methods for Spheres and SO(3), 2004. 51

[105] K. Yoshikawa, Y. Nakata, K. Yamada, and M. Nakagawa. Early pathological changes in the parkinsonian brain demonstrated by diffusion tensor MRI. *J. Neurol. Neurosurg. Psychiatry*, 75:481–484, 2004. 1, 3, 24

[106] C. S. Yu, K. C. Li, Y. Xuan, X. M. Ji, and W. Qin. Diffusion tensor tractography in patients with cerebral tumors: A helpful technique for neurosurgical planning and postoperative assessment. *European Journal of Radiology*, 56(2):197–204, 2005. 1, 24

[107] S. Zaroubi and G. Goelman. Complex denoising of MR data via wavelet analysis: application for functional MRI. *Magnetic Resonance Imaging*, 18(1):59–68, January 2000. 8