Variable Splitting as a Key to Efficient Image Reconstruction

by

Sudipto Dolui

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2012

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

The problem of reconstruction of digital images from their degraded measurements has always been a problem of central importance in numerous applications of imaging sciences. In real life, acquired imaging data is typically contaminated by various types of degradation phenomena which are usually related to the imperfections of image acquisition devices and/or environmental effects. Accordingly, given the degraded measurements of an image of interest, the fundamental goal of image reconstruction is to recover its close approximation, thereby "reversing" the effect of image degradation. Moreover, the massive production and proliferation of digital data across different fields of applied sciences creates the need for methods of image restoration which would be both accurate and computationally efficient. Developing such methods, however, has never been a trivial task, as improving the accuracy of image reconstruction is generally achieved at the expense of an elevated computational burden. Accordingly, the main goal of this thesis has been to develop an analytical framework which allows one to tackle a wide scope of image reconstruction problems in a computationally efficient manner. To this end, we generalize the concept of variable splitting, as a tool for simplifying complex reconstruction problems through their replacement by a sequence of simpler and therefore easily solvable ones. Moreover, we consider two different types of variable splitting and demonstrate their connection to a number of existing approaches which are currently used to solve various inverse problems. In particular, we refer to the first type of variable splitting as Bregman Type Splitting (BTS) and demonstrate its applicability to the solution of complex reconstruction problems with composite, cross-domain constraints. As specific applications of practical importance, we consider the problem of reconstruction of diffusion MRI signals from sub-critically sampled, incomplete data as well as the problem of blind deconvolution of medical ultrasound images. Further, we refer to the second type of variable splitting as Fuzzy Clustering Splitting (FCS) and show its application to the problem of image denoising. Specifically, we demonstrate how this splitting technique allows us to generalize the concept of neighbourhood operation as well as to derive a unifying approach to denoising of imaging data under a variety of different noise scenarios.

Acknowledgements

I would like to express my sincere gratitude to Professor Oleg Michailovich, my thesis advisor, for his unwavering support throughout this journey. Collaborating with him has been a tremendous learning experience which has gone a long way in helping me become an independent researcher.

I am thankful to my committee members Professor Fakhri Karray, Professor Patrick Mitran, Professor John Zelek and Professor Greg Stanisz for providing invaluable suggestions to improve my thesis.

I would like to thank Professor Ravi Mazumdar and Professor Edward Vrscay for providing excellent insights through their courses that proved greatly valuable for my research.

I must also mention my gratitude to the University of Waterloo, in general, and the Department of Electrical and Computer Engineering, in particular, for providing me the opportunity and the infrastructure to pursue my graduate studies at such a world-class academic environment.

I would also like to thank all my friends and labmates whose moral support had a huge impact on keeping me going. Finally, I thank my parents, Mina and Ajit, for being my pillars of strength throughout this journey. I bow down to their immense patience in not only putting up with such an ill-behaved son, but being there every step of the way.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADMM	Alternating Direction Method of Multipliers
AWGN	Additive White Gaussian Noise
BTS	Bregman Type Splitting
CS	Compressed Sensing
CRSM	Correlative Rational Similarity Measure
CSSM	Correlative Subtractive Similarity Measure
dMRI	Diffusion Magnetic Resonance Imaging
DTI	Diffusion Tensor Imaging
FCS	Fuzzy Clustering Splitting
FISTA	Fast Iterative Shrinkage-Thresholding Algorithm
HARDI	High Angular Resolution Diffusion Imaging
MAP	Maximum A Posteriori
ML	Maximum Likelihood
NCCS	Non-Central Chi Square
NLM	Non-Local Means
NMSE	Normalized Mean Square Error
MRI	Magnetic Resonance Imaging
ODF	Orientation Distribution Function
PSF	Point Spread Function
RIP	Restricted Isometry Property
RSM	Rational Similarity Measure
SD	Sparse Deconvolution
SIF	Sparse Inverse Filtering
SM	Similarity Measure
SNR	Signal to Noise ratio
SSIM	Structural SIMilarity
SSM	Subtractive Similarity Measure
TV	Total Variation
WML	Weighted Maximum Likelihood
WSN	Wireless Sensor Network

Chapter 1

Introduction

Imaging is a broad field which encompasses a large number of areas that include development of the imaging protocol, generation of the images, their modification, visualization and finally their analysis [1, 2]. Consequently, it is a highly interdisciplinary field involving physicists, mathematicians, engineers and biologists to name a few. Among the different fields of imaging sciences, two very important fields are image reconstruction and image enhancement [1, 2, 3, 4, 5, 6]. Although both of these fields aim at recovering a close approximation of the original image, their difference lays in the fact that image enhancement inputs an image and improves its quality, whereas image-reconstruction reconstructs an image from imagery data. For example, the process of generation of an ultrasound image from the acoustic waves reflected from different organs within a human body is considered to be a problem of image reconstruction whereas increasing the sharpness or removing noise from an image is considered to be image enhancement. In this thesis, we will use the term image reconstruction (aka image restoration) to refer to both; i.e. the term will be used in a generic sense to denote the broad concept of recovering an approximate version of the image of interest.

The process of image reconstruction is considered to be of great importance in the field of imaging science as a standalone procedure as well as a preprocessing step in several fields of image processing and computer vision [7, 8]. The principles of image reconstruction are based on the formal theory of inverse problem which we are going to discuss in this chapter. However, let us first discuss the image-acquisition model which will be used throughout this thesis.

1.1 Image Acquisition Model

Image acquisition is performed by collecting imagery data containing full or partial information to reconstruct the image [1]. Irrespective of the type of imaging modality, the acquired data is frequently contaminated with noise [1, 9, 10, 11, 12, 13, 5] which is defined to be a random, unwanted signal that does not convey any meaningful information. For example, in the case of a charge-coupled device (CCD) sensor, the noise is the result of inherent natural variation of the incident photon flux, imperfection of the acquisition device as well as environmental disturbance. It should be noted that, besides being an effect of device limitation or environmental disturbance, noise can also be a result of the physics of the image acquisition process itself. As an example, one can consider coherent imagery (e.g. synthetic aperture radar, ultrasound imaging) [14, 15] where the acquired image is contaminated with noise resulting from the interference of electromagnetic waves travelling via different paths and thus having different path lengths. In general, depending on the principles of the acquisition process as well as physical or technological limitations, the contaminating noise can follow different statistical distributions, important examples of which are Gaussian, Poisson and Rician.

In addition to the problem of *inaccuracy* as a result of the contaminating noise, the acquired data is often *incomplete* which is associated with a "loss" of informational content of the image. For example, the acquired images are often blurred as compared to the original image due to loss of high frequency components of the image. This can happen because of motion blur, camera defocus or imperfections inherent to the acquisition device/principle. In addition, incompleteness can also result from acquisition of a smaller amount of data because of device limitation or other physical restrictions such as limited acquisition time.

In this thesis, we are going to consider image acquisition models that can be classified as a linear measurement model. To specify the model, let us define $X \in \mathbb{R}^N$ and $Y \in \mathbb{R}^M$, with $M \leq N$, to be the original and the observed data respectively. It should be noted that although the above data representation considers one independent variable (e.g. time or one dimensional space), data for multiple independent variables can be vectorized by arranging it in a lexicographic order. As a next step, let us define a linear operator $\mathcal{A} \in \mathbb{R}^{M \times N}$ which describes the measurement model and can be associated to the incompleteness of the data. Formally,

$$\mathcal{A} : \mathbb{R}^N \mapsto \mathbb{R}^M : \; X \mapsto Z = \mathcal{A}X. \tag{1.1}$$

Moreover, since the acquired data is contaminated with noise, we define an additional operator $\mathcal{N} : \mathbb{R}^M \mapsto \mathbb{R}^M$ which describes the inaccuracy in the data. Formally,

$$Y = \mathcal{N}\{Z\}. \tag{1.2}$$

In connection to the previous discussion, the operator $\mathcal{A}$ can model a blurring artifact (Toeplitz matrix) and/or subsampling operation. On the other hand, the operator $\mathcal{N}$ can model contamination with noise following, for example, Gaussian, Poisson or Rician statistics. Finally, the combined data acquisition model is given by

$$\mathcal{N} \circ \mathcal{A} : \mathbb{R}^N \mapsto \mathbb{R}^M : \; X \mapsto Y = \mathcal{N}\{\mathcal{A}X\}. \tag{1.3}$$

The different types of $\mathcal{N}$ and $\mathcal{A}$ that have been used in this thesis are summarized in Table 1.1.

Table 1.1: The types of $\mathcal{N}$ and $\mathcal{A}$ operators that have been used in this thesis

	Noise Type ($\mathcal{N}$)			
$\mathcal{A}$	Gaussian	Rician	Poisson	Multiplicative Gamma and Rayleigh
Identity	✓	✓	✓	✓
Convolution	✓			
Subsampling	✓			

As mentioned before, the procedure of image reconstruction is based on the formal theory of inverse problem which is discussed next.

1.2 Inverse Problems

The concept of "inverse problem" [16, 2] can be well described by the statement of J. B. Keller [17]. "We call two problems inverses of one another if the formulation of each involves all or part of the solution of the other. Often, for historical reasons, one of the two problems has been studied extensively for some time, while the other is newer and not so well understood. In such cases, the former problem is called the *direct problem*, while the latter is called the *inverse problem*." In essence, the forward problem and the inverse problem are related by a sort of *duality* in the sense that one problem can be obtained from the other by exchanging the role of the data and the unknown variable. In real-life, inverse problem is a general framework used to extract the information about the unobserved object of interest from observed measurements. They are common in many branches of science that include classical mechanics, scattering and diffraction theory and wave propagation theory [16].

In the field of imaging science, the formation of the observed image from the original image given the characteristic of the instruments is referred to as the direct problem. With

reference to the previous section, (1.1) models the direct problem and the operator $\mathcal{A}$ is referred to as the *forward operator* [18]. The corresponding inverse problem is the problem of reconstructing the original image by inverting the effect of $\mathcal{A}$. It should be noted that the data is always contaminated with noise and hence the inversion of the operator $\mathcal{A}$ has to be performed using the noisy data Y given in (1.3). Formally, the inverse problem is to obtain a function g such that

$$\hat{X} = g(Y), \tag{1.4}$$

where $\hat{X}$ is an approximation of the quantity of interest. In this connection, the inverse problem can also be interpreted as to obtain a solution $\hat{X}$ so that upon its substitution, the forward problem (1.1) is satisfied up to a reasonably small approximation error.

The direct problem, as modelled by $\mathcal{A}$, is associated with a loss of information. For instance, the operation of blurring of the original image is associated with the loss of its high frequency content. Hence the solutions of these inverse problems need to be associated with an increase in the informational content of the image. This is associated with the concept of *ill-posedness* of the inverse problem which will be addressed in the next section.

Depending on the type of $\mathcal{A}$, the inverse problem can be of different types and in this thesis we will deal with the following three types.
Type 1: In the case when $\mathcal{A} = I$ i.e. an identity matrix, the inverse problem is only concerned with *denoising* or removal of noise [5, 11, 12, 9].
Type 2: $\mathcal{A}$ can be a Toeplitz matrix which corresponds to a filtering operation using a linear shift invariant convolution kernel [3, 19, 20, 21, 22]. In this thesis, we are interested in the particular case where the filtering operation results in blurring the image. Accordingly, the inverse problem is that of deblurring or deconvolution. The process of deblurring can be blind or non-blind depending on whether the knowledge of the forward operator (or, equivalently, its associated *point spread function*) is available or not. In the latter case, the process of identification of the blur adds an extra level of difficulty to the process of the inverse problem.
Type 3: A different scenario is when $\mathcal{A}$ corresponds to subsampling operation. The corresponding inverse problem is known as compressed sensing. In this case, the problem is to recover $X \in \mathbb{R}^N$ from $Y \in \mathbb{R}^M$ with $M \ll N$, i.e. from fewer samples as compared to that of the dimension of the original signal [23, 24, 25, 26].

The three types of inverse problem mentioned above are summarized in Table 1.2.

Table 1.2: The type of inverse problem corresponding to the operator $\mathcal{A}$

$\mathcal{A}$	*Identity*	*Convolution*	*Subsampling*
Inverse Problem	*Denoising*	*Deblurring*	*Compressed Sensing*

1.3 Ill-Posedness and *A Priori* Information

The mathematical definition of well-posedness is provided by the French mathematician Jacques Hadamard. Formally, a problem is well-posed if [27]

1. there exists a solution to the problem (existence),
2. there is at most one solution to the problem (uniqueness),
3. the solution depends continuously on the data, i.e. it should not vary considerably for small variation of the data (stability).

A problem which does not satisfy one or all of the above conditions is said to be *ill-posed* (aka incorrectly posed or improperly posed). Unfortunately, most of the inverse problems that we deal with are ill-posed. For instance, image blurring is associated with attenuation or complete suppression of high frequency content and hence the inverse problem of recovering the same is associated with non-uniqueness and/or instability of the solution.

Mathematically, ill-posedness is associated with reduced rank of the operator $\mathcal{A}$ with respect to the dimension of the original image. As a result, the solution obtained by *naive inversion* of the direct operator becomes non-unique. In addition to that, since the observed data is contaminated with noise, direct inversion of the ill-posed matrix $\mathcal{A}$ results in amplification of the noise. Consequently, the resulting solution completely obscures the physical solution corresponding to the noise-free data.

The problem of non-uniqueness or instability can be addressed by restricting the set of admissible solutions by using some additional information (often called *a-priori* or *prior* information) on the solution. It should be noted that the additional information cannot be derived from the observed data or the properties of the mapping operator $\mathcal{A}$, but is obtained from the known physical properties of the original image. Example of additional information can include boundedness and/or non-negativity of the original image intensities. As an additional example, one can assume natural images to be piecewise smooth. This is incorporated in the reconstruction procedure by imposing the constraint that the image gradient is small with occasional large values. Additionally, a priori information

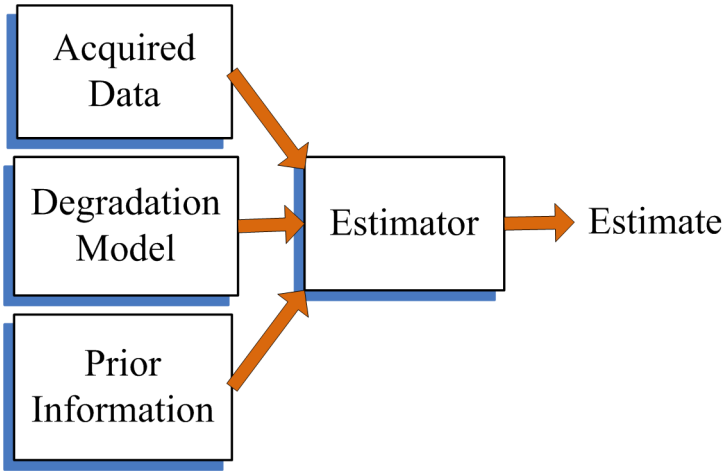


Figure 1.1: Reconstruction approach

can also be imposed on the image in some transform domain. For example, in the domain of wavelet transform, the coefficients can be assumed to be sparse [11]. Although the a priori information can solve the problem of ill-posedness of the forward operator, one should be careful in choosing the prior because an inappropriately chosen prior will result in an undesirable solution. The a priori information is incorporated in the reconstruction problem along with the constraint that the solution conforms to a predefined measurement model (data fidelity). The effects of the two components are combined to formulate an *optimization problem* whose solution results in the desired estimate.

Statistical properties of the original image can also serve as additional information to recover an image. In this case, it is assumed that the image to be recovered is a realization of a random process with known probability distribution which can be obtained through previous experience about the possible image to be restored. This is normally incorporated in a formal *Bayesian framework*. Although derived using different philosophies, the Bayesian and the optimization frameworks can be formally connected and we are going to discuss this in the next section. A block diagram, showing the reconstruction process using the data fidelity term and the prior information is shown in Fig. 1.1.

1.4 Bayesian Framework

The Bayesian framework provides a formal approach to formulate a reconstruction problem which incorporates the noise statistics and the a priori information about the object

of interest. The theory is based on the *Bayes' theorem* which combines the statistical distribution of the contaminating noise with the prior probability density function of the unobserved image to express the *posterior probability density* of the latter given the noisy data. To give a formal representation, let $X \in \mathbb{R}^N$ represents the original image and $Y \in \mathbb{R}^M$ represents the observed data. The noise statistics is incorporated in the conditional probability density $p_{Y|X}(y \mid x)$ whereas $P_X(x)$ denotes the prior probability density of X[1]. Subsequently, the posterior probability density of X having observed the noisy data Y is given by

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y \mid x)p_X(x)}{p_Y(y)}. \tag{1.5}$$

Based on different optimality conditions, one may obtain several estimates using the posterior probability density. Among these, of particular importance is the *maximum a-posteriori* (MAP) estimate which is defined as the mode of the posterior density. Formally

$$\hat{X}_{MAP} = \arg\max_X p_{X|Y}(x \mid y). \tag{1.6}$$

Unfortunately, the MAP estimate admits analytical solutions only in a limited number of cases and one needs to rely on solving optimization methods to obtain the solution. To show the connection between the Bayesian and the optimization frameworks, we first note that a lot of conditional or prior densities fall in the category of the exponential densities. Consequently, it is often computationally advantageous to maximize the log of the posterior density function instead of the actual function as given by

$$\hat{X}_{MAP} = \arg\max_x \{\log(p_{Y|X}(y \mid x)) + \log(p_X(x))\}, \tag{1.7}$$

and this is often reformulated as

$$\hat{X}_{MAP} = \arg\min_x \{-\log(p_{Y|X}(y \mid x)) - \log(p_X(x))\}. \tag{1.8}$$

In order to show the relation between the Bayesian framework and the optimization methods, let us consider a simple and widely used model where $\mathcal{N}$ is additive white Gaussian noise (AWGN). More specifically, the observed data and the original image are related by

$$Y = \mathcal{A}X + \eta, \tag{1.9}$$

where the elements of $\eta \in \mathbb{R}^M$ follow independent and identically distributed (i.i.d.) AWGN model. In such a case, the conditional density function of Y, given X, is given by

$$p_{Y|X}(y \mid x) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left\{-\frac{\|y - \mathcal{A}x\|_2^2}{2\sigma^2}\right\}, \tag{1.10}$$

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

where the $\|\cdot\|_2$ denotes the ℓ_2 norm defined as

$$\|z\|_2 = \sqrt{\sum_{i=1}^{M} z_i^2}, \quad z \in \mathbb{R}^M. \tag{1.11}$$

On the other hand, let the prior density be given by

$$p_X(x) = \frac{1}{Z} \exp\left\{-\frac{\phi(x)}{\mu}\right\}, \tag{1.12}$$

where $\phi(x) : \mathbb{R}^N \mapsto \mathbb{R}$ and Z is a normalizing constant. In such a case, the MAP estimate $\hat{X}_{MAP}$ can be written as

$$\hat{X}_{MAP} = \arg\min_x \left\{\frac{1}{2}\|y - \mathcal{A}x\|_2^2 + \lambda\phi(x)\right\}, \tag{1.13}$$

where $\lambda = \sigma^2/\mu$. It should be noted that the above expression can be derived from a different perspective. In particular, the first term in (1.13) reflects the fact that the intensities of the reconstructed image should be close to those of the observed noisy image. Likewise, the second term, namely $\phi(x)$, acts as a regularizing function which regularizes the solution based on the assumption of the original image. In this case, $\lambda > 0$ in (1.13), commonly known as regularization constant, maintains a balance between the data-fidelity term and the regularizer. It should be noted that one can add multiple regularization terms with associated constants. Also, one can obtain different data fidelity terms resulting from other types of contaminating noise. The expression on the right hand side of (1.13) that needs to be minimized to obtain the estimate of the original image is often referred to as the *cost function* [28].

It should be noted that depending on the size of optimization vector or the complexity and the number of regularization terms, the above mentioned optimization problem can result in computationally involved optimization problems as explained in the next section.

1.5 Optimization and Complexity

As mentioned above, the MAP estimator does not admit a closed form solution in general and one needs to use optimization techniques to find the estimate of the quantity of interest [29, 4, 22, 30]. As a result, the computational burden increases with increased complexity of the cost function [31, 32, 33]. The increase in complexity of the cost function may happen

for a number of reasons. In general, the a priori probability of the original image of interest should account for all kinds of dependencies between all the pixels of the image. Similarly, the degradation model should also consider the dependencies between the noise intensities contaminating the different pixels. These complexities couple with the fact that one can include multiple regularizers in the reconstruction process in order to obtain a better estimate. Hence the optimization problem is hard to solve for the case of a large number of variables which, however, happens to be the case in the field of image processing [34]. Furthermore, the increase in computational complexity is associated with the increase in time complexity as well. Consequently, it would not be possible to implement the algorithm in real time. As a result, the design of an algorithm often trades off complexity in exchange of performance.

A possible approach to obtain a computationally tractable solution is to simplify the optimization problem by replacing it with a sub-optimal problem. Such a simplification can be obtained by ignoring the interdependencies between the intensities of different pixels or that between the contaminating noises. Simplification can also be attained by considering dependencies only between the neighbouring pixels as opposed to the whole image or by performing asymptotic simplification of the cost function. Obviously, such a solution will not be optimal, but is often preferred because of the reduction in the computational complexity, provided the performance is acceptable for the particular application in hand.

In the next section, we introduce one such methodology which has the potential of reducing the computational complexity in several applications while providing reasonable quality of reconstruction.

1.6 Variable Splitting

In this thesis, we extend the concept and applicability of the idea of variable splitting as a methodology to simplify complex estimation problems. In the field of image restoration, the phrase "variable splitting" has been used in the context of solving optimization problems [32, 30] to obtain an estimate of the quantity of interest. Specifically, it involves replacing an optimization problem by an equivalent one through the introduction of new variables with the end goal of reducing the computational complexity.

In this thesis, we use the term "variable splitting" in a more generic sense. In particular, we use it to replace the problem of estimating a variable of interest by a sequence of sub-problems through the introduction of additional variables. Next, the additional variables are estimated by solving the sub-problems and these estimates, in turn, are used to estimate

the original quantity of interest. The reduction in the computational complexity is obtained because of the fact that the sub-problems are simpler to solve as compared to the original problem and their solutions can be obtained by existing efficient computational means. We demonstrate the utility of this method by solving some important computationally complex problems.

The thesis focuses on two types of variable splitting techniques which are presented next.

1.6.1 Bregman Type Splitting (BTS)

The Bregman type splitting (BTS), popularly known as split Bregman technique [33], is a method of superseding a complex convex optimization problem by a sequence of simpler sub-problems. The splitting technique is equivalent to the *alternating direction method of multipliers* (ADMM)[35, 36, 34, 31] and these two terms will be used interchangeably in this thesis. The general form of the optimization problem that can be solved using BTS is given by

$$\min_{x} H(x) + \sum_{i=1}^{Q} G_i(A_i x + b_i). \tag{1.14}$$

Here $H : \mathbb{R}^N \mapsto \mathbb{R}$ and $G_i : \mathbb{R}^{n_i} \mapsto \mathbb{R}, i = 1 \dots Q$ are closed proper convex functions (defined later in this thesis), $A_i \in \mathbb{R}^{n_i \times N}$, $b_i \in \mathbb{R}^{n_i}$ for $i = 1 \dots Q$. It can be clearly seen that the optimization variable x is *coupled* with several other variables and direct solution of the optimization problem, when considered as a whole, can be non-trivial. However, a simplification can be obtained by replacing the original problem in (4.14) by a sequence of sub-problems. Specifically, the problem in (4.14) can be converted into an equivalent constrained minimization problem given by

$$\begin{aligned} &\min_{x,u} H(x) + F(z) \\ &\text{s.t. } z = Ax + b \end{aligned} \tag{1.15}$$

where

$$F(z) = \sum_{i=1}^{N} G_i(z_i), \quad z = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}, \quad A = \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}, \tag{1.16}$$

The goal of this splitting method is to produce algorithms consisting of simple, easy-to-compute steps that can deal with each term of the cost function one at a time. Specifically,

using an augmented Lagrangian formulation, one can solve the problem using the following steps

$$(x^{k+1}, z^{k+1}) \in \arg\min_{x,z} H(x) + F(z) + \frac{\mu}{2}\|Ax + b - z - d^k\|_2^2 \tag{1.17}$$

$$d^{k+1} = d^k - (Ax^{k+1} + b - z^{k+1}) \tag{1.18}$$

It can be shown that the sequence of iterates $\{x^k, z^k\}$ converges to the primal problem in (1.15) and the sequence of iterates $\{d^k\}$ converges to the solution of the Lagrange dual problem of (1.15) [31].

The minimization problem in (1.17) is in general non-trivial. A possible way to solve (1.17) is to use a non-linear block-Gauss-Seidel technique which alternately minimizes the cost function with respect to x and z (while keeping the other variable fixed) until convergence. However, experimental evidence [33] as well as theoretical guarantee [34, 37] exist that one can perform one step of the non-linear block Gauss-Seidel method followed by the update (1.18) to achieve stable convergence. In other words, the minimization problem in (1.17) can be replaced by

$$x^{k+1} \in \arg\min_{x} H(x) + \frac{\mu}{2}\|Ax + b - z^k - d^k\|_2^2 \tag{1.19}$$

$$z^{k+1} \in \arg\min_{z} F(z) + \frac{\mu}{2}\|Ax^{k+1} + b - z - d^k\|_2^2 \tag{1.20}$$

Because of the separable structure of (1.20), it is possible to split it further to result in

$$z_i^{k+1} \in \arg\min_{z_i} F(z_i) + \frac{\mu}{2}\|A_i x^{k+1} + b_i - z_i - d_i^k\|_2^2 \tag{1.21}$$

The solution of each of these problems for most of the functions can be obtained through simple techniques. Moreover, it can be shown that the solution of the decoupled problem is same as that of the original problem in (4.14) [34, 37]. Consequently, splitting reduces the overall computational complexity while being able to obtain the same solution.

1.6.2 Fuzzy Clustering Splitting (FCS)

In this thesis, we introduce the formal concept of Fuzzy Clustering Splitting (FCS) as a generic approach to the problem of image de-noising. In order to present the method, we note that the degraded image can be considered to be a realization of a random process parametrized by the intensities of the underlying noise free image. The random process can

be characterized by a joint probability density function which accounts for the statistical distribution of the pixel intensities as well as the complex dependencies that may exist between them. Needless to say, the problem of inference of the quantity of interest considering these dependencies is non-trivial. Consequently, certain reasonable simplifications of the original model are required to render the reconstruction problem practical.

A particular type of simplification can be obtained based on the characteristics of self similarity of natural images [9, 10]. Specifically, in the case of natural images, the structures, patterns and the pixel values tend to repeat themselves. The repetition of the pixel intensities can simplify the statistical model which can now be looked upon as a mixture of ergodic processes. Each ergodic process is characterized by a single parameter which corresponds to the same noise-free pixel intensity corresponding to the locations where the intensity is repeated. Consequently, if the pixels can be clustered based on the ergodic processes they belong to, the underlying parameter corresponding to each process can be estimated based on the noisy realizations.

One problem with using such an approach stems from the fact that the prior information of the location of the pixels belonging to a particular cluster is not available. A naive approach of obtaining such a classification would be to assume that the neighbouring pixels belong to the same ergodic process. However, such an assumption leads to blurring artifacts at the regions containing edges or fine structures where the above mentioned assumption fails. A better approach is to perform the clustering adaptively. Accordingly, the denoising procedure can be divided into two stages, viz.

1. identify the pixels whose intensities belong to the same ergodic process, and
2. estimate the parameter of the process based on the pixel intensities within the same cluster.

The first step of the above procedure can be performed by considering a finite number of clusters and subsequently assigning each pixel to one of these clusters. However, such an approach, in general, requires a priori knowledge of the number of clusters. Moreover, the procedure of assigning pixels to each cluster is prone to error because it is performed based on the noisy data. An alternative is to use fuzzy clustering by considering an ergodic process corresponding to each pixel and assigning the other pixels to the cluster with certain probability. Thereafter, these pixels are assigned weights which are proportional to their probability of being part of the cluster. These weights can be considered to be auxiliary variables which are artificially introduced in order to simplify the reconstruction problem. These variables have to be estimated for each pixel and subsequently they are used to

obtain an estimate of the original quantity of interest. Since the artificially introduced variables relate to the fuzzy clustering method, we refer to this type of variable splitting technique as the *fuzzy clustering splitting* (FCS) technique. Important filters like range filters [1, 38], bilateral filter [39], and the recently developed non-local means filter [9, 10] can be classified as denoising methods using FCS.

1.7 Significance of Variable Splitting

In real-life application of reconstruction of an image, obtaining an optimal solution considering all possible factors is often not possible due to several practical considerations. Computational resource constraint is often the most important practical consideration. Consequently, several applications often demand decrease in computational complexity in exchange of reasonable reduction in quality of reconstruction. The principles of variable splitting provide a novel framework to address the above problem. Specifically, our goal is to obtain significant reduction in computational complexity while attaining close to optimal performance. In this thesis we consider two types of variable splitting techniques. The practical value of the proposed methodologies will be demonstrated using a number of imaging modalities throughout this thesis.

1.8 Contributions of the Thesis

The goal of the thesis is to expand the spectrum of reconstruction techniques in the field of imaging science using the principles of variable splitting. In particular, we introduce a number of novel approaches for solving a large number of problems in modern day imaging sciences. Each of these approaches are specifically addressed to maintain a balance between the reconstruction accuracy and computational complexity. In that sense, the goal of the thesis is to provide practical algorithms capable of solving real life applications as opposed to providing optimal performance to a specific problem. We also provide important theoretical connections between important concepts originating from different fields of image reconstruction.

The contribution of the thesis can be broadly divided into two parts based on the two types of splitting discussed above.

1.8.1 Formulation of BTS for Image Reconstruction

In the first part of the thesis, we use BTS to address optimization problems with composite constraints where the constraints arise from different domains of definition of the image. In absence of the splitting scheme, solutions of such problems are computationally complex and often intractable. We show how the applications of BTS allow efficient implementation of each problem. In this thesis, we address two different classes of problems whose efficient solutions are possible using BTS. Subsequently, we demonstrate the practical values of the methods by addressing two important problems in the field of medical imaging.

The first of the two contributions presents a new class of problems where the signals are multidimensional vector valued and are associated with the requirement of acquisition and processing of large amounts of data. As a solution, we propose to acquire reduced amount of data and subsequently use the tool of compressed sensing to reconstruct the whole signal. This is facilitated by solving an optimization problem which couples regularizations defined over multiple domains of definition of the signal. The computational complexity has been simplified using BTS. As an example, we propose a solution to the problem of prohibitively long acquisition time inherent in diffusion magnetic resonance imaging (MRI). Specifically, the problem has been solved by acquiring fewer samples compared to what is required for efficient reconstruction, while compensating for the undersampling artifacts using the theory of compressed sensing. Diffusion MRI is a very important imaging technology that can detect a number of neurological disorders which are not identifiable by conventional MRI. Thus for example, it has the potential of early detection of ischemic stroke [40] which accounts for 88 percent of all strokes. However, the problem of long acquisition times in diffusion MRI is a serious issue which prohibits its extensive use in several clinical settings. Consequently, the proposed solution for reducing the acquisition time can have serious impact in clinical MRI applications.

As an additional contribution, we propose an efficient solution to the problem of blind deconvolution using the "hybrid" deconvolution approach of [41]. In particular, the approach requires solving an optimization problem which couples regularizations defined over the spatial as well as the frequency domains. The solution of the optimization problem is non-trivial because of such coupling of the regularization terms. However, the optimization problem has been decoupled using BTS which leads to a computationally efficient solution. As an application of the technique, we address the problem of increasing the resolution and contrast of medical ultrasound images which are characterized by low resolution and reduced contrast.

1.8.2 Formulation of FCS for General Imaging

In the second part of the thesis, we use the concept of FCS to denoise images contaminated with different types of measurement noise. In particular, we propose a unifying denoising approach that can be applied to a range of different noises in a consistent and accurate manner. The acquisition process of many imaging techniques has a trade-off between acquisition time, resolution, and the signal to noise ratio (SNR). In many cases, in order to reduce the acquisition time, or to obtain better resolution, the acquired images are often noisy. Consequently, the only method of removing the contaminating noise is through the use of post-processing techniques. In this thesis, we consider the cases of Poisson, Rician, multiplicative Gamma and Rayleigh noise. The types of noises mentioned above are prevalent in a number of imaging modules and hence is of significant importance in the field of imaging science.

1.9 Organization of the Thesis

The thesis presents a number of approaches relevant to the field of image reconstruction. Accordingly, it has been divided into several chapters based on the presentation of each of these approaches. The content of the thesis has been organized as follows.

Chapter 2 provides a survey of the existing literature on the different frameworks and approaches that are relevant to our research. Chapter 3 provides the necessary technical details that will be necessary in the presentation of our research. The contributions of the thesis have been presented in Chapters 4, 5 and 6. In particular, Chapter 4 presents the compressed sensing framework along with its application to the field of diffusion MRI. Likewise, Chapter 5 presents the computationally efficient approach for solving the problem of hybrid blind deconvolution while showing its application to deconvolve ultrasound images. Both chapters demonstrate the application of BTS as a methodology to simplify computations. On the other hand, Chapter 6 demonstrates the importance and efficiency of FCS in denoising images contaminated with different types of noise. The results of each of the proposed methodologies have been presented at the end of the corresponding chapter. Finally, the thesis has been concluded in Chapter 7 with a summary of the salient points leading towards directions for future research.

Chapter 2

Literature Review

2.1 Inverse Filtering using Bayesian Methods

In the context of image processing, inverse filtering is a highly developed field. In general, inverse filtering is performed through the use of a priori information and, as mentioned in the previous chapter, this can be done using the formal theory of a Bayesian framework. In this section, we are going to refer to the important works in the field of inverse filtering that are relevant to our research.

2.1.1 Image Denoising

Image denoising is a very important post processing technique in the field of imaging science [1]. In many cases, the contaminating noise obscures important details of the image and makes subsequent image processing tasks difficult. The objective of image denoising is to remove the contaminating noise without removing significant structures within the image or without producing disturbing artifacts.

The current arsenal of image denoising methodologies is immense, which makes their objective comparison and classification a very non-trivial task. For this reason, we are going to discuss the following three important groups of denoising techniques, which encapsulate a number of important approaches. Specifically, the first group encompasses variational techniques, in which the denoised image is obtained as a solution to a partial differential equation (PDE) [5, 13, 42, 43, 44, 45, 46, 47]. Typically, the PDE is solved numerically and it represents a gradient flow corresponding to minimization of a certain cost function,

whose definition is problem dependant. Thus, for example, the seminal work [5] introduced the concept of total variation denoising, which finds a solution of minimal total variation seminorm that conforms with a predefined measurement model.

A different group of denoising methods can be characterized as transform-based techniques which exploit diagonal operators applied in the domain of sparsifying linear transforms [48, 12, 49, 50, 51, 52, 53, 11, 54]. Many methods of this type derive from the seminal works of D. Donoho on wavelet shrinkage, as originally reported in [53, 11, 54], while being different in the way the transformation (e.g., wavelet) coefficients of the original image are modelled. Thus, for example, in [12] the wavelet coefficients are assumed to follow a Gaussian scale mixture model, which allows their recovery using a Bayesian estimation framework. Using a different line of arguments, the work in [49] proposed a robust and adaptive denoising method which can adapt to various types of measurement noise.

The third group of image denoising methods exploits the concept of neighbourhood filtering [38, 10, 55], where a particular pixel intensity is estimated as a weighted sum of its neighbouring pixel intensities. The simplest of such methods uses non-adaptive and spatially-invariant weights [1] where the weights in general decrease with increased spatial distance. Unfortunately, the blurring artifact introduced by such filters is known to be their most critical drawback, which can be alleviated by setting the weights adaptively to image intensities, as it is done in range [38] and bilateral [56, 57, 39, 58, 59, 60, 61, 62] filtering. In particular, in the case of range filter, the weights are dependent on the similarities of the pixel intensities. On the other hand, the bilateral filter computes the weights based on the similarity of the intensities as well as the proximity of their spatial locations. However, the weights in the case of range or bilateral filters are reliable only for a mild to moderate level of noise. Yet even more accurate reconstruction results are obtainable with *non-local means* (NLM) filtering, which has been shown to outperform a wide range of alternative denoising techniques for the case of additive white Gaussian (AWG) noise contamination [9, 10]. The weights in the non-local means filters are computed based on the similarity of the patches centred around the pixels. Thus it can be considered an improved version of the range filter because the weights for the NLM filter are less susceptible to noise as compared to the range filter. A number of improvements to the above filter have later been proposed in [63, 64].

The success of NLM filtering has motivated other researchers to apply the method to several applications [65, 66, 67]. Moreover, several theoretical frameworks have been developed to incorporate various types of noise statistics [68, 7, 69]. For example, in the work of [68], a non-local means filter has been derived using a Bayesian approach which can incorporate other types of noise statistics. On the other hand, the work in [7] proposes an iterative weighted maximum likelihood (WML) scheme to denoise the images.

In this work, the similarity between pixel intensities is computed based on formal statistical consideration. It has been shown in [69], however, that the similarity measure in [7] yields suboptimal results in the case of multiplicative noises. Consequently, to alleviate this problem, a different similarity measure was proposed in [69].

Contribution 1: In this thesis, we will propose a unifying framework for the definition of weights in the case of non-local means filtering. This approach is capable of overcoming some of the drawbacks of the previous methods. Specifically, the previous approaches for computing similarity measures cannot be regarded as general, as it has been found that the NLM method which is optimal for a particular type of noise is likely to produce inconsistent denoising results for a different type of noise. The approach that will be presented in this thesis, on the other hand, is devoid of these drawbacks and is guaranteed to produce consistent results irrespective of the type of noise.

2.1.2 Image Deconvolution

In many practical scenarios of image acquisition, the acquired images are often blurred and are of low contrast because of suppression or complete attenuation of high frequency components [19, 70, 71, 72, 73, 74]. This can happen because of the low bandwidth of the acquisition device, relative motion between the acquisition device and the subject, atmospheric turbulence and several other factors. The field of deblurring or deconvolution aims at recovering a sharper version of the acquired image by inverting the effect of the blurring operator. The name deconvolution stems from the fact that the blurring artifact is often modelled by a convolution operation and hence its inverse process is termed as deconvolution. Deconvolution is important in many applications such as astronomical imaging [75, 76], remote sensing [77] and medical imaging [78, 79, 80, 41].

In general, the process of image deconvolution can be either non-blind or blind. In the former case, an estimate of the blurring convolution kernel, also referred to as point spread function (PSF), is assumed to be known prior to the deconvolution process. This category of deconvolution methods includes a wide variety of techniques; for example, inverse filtering, Wiener filtering, least-square filtering, recursive Kalman filtering, and iterative deconvolution methods [3, 81, 82, 83, 84, 4, 29, 22].

Unfortunately, in many practical applications, the prior knowledge of the PSF is not available. For example, in applications like astronomy or remote sensing, it is difficult to obtain a priori information about the scene or the type of degradation that has never been imaged before [75, 77]. Moreover, in applications like turbulent imaging, it is very difficult to characterize the fluctuation of the PSF [76]. In these situations, both the

PSF and the quantity of interest have to be estimated from the degraded data and the deconvolution techniques are referred to as blind [19]. Blind deconvolution can be classified into different types depending on the method of estimating the PSF. For example, the PSF can be estimated separately prior to the operation of deconvolving the image and subsequently a non-blind procedure can be used to perform the deconvolution [78, 79, 85, 20, 80, 86]. On the contrary, a different type of blind deconvolution algorithms estimates the PSF and the original image concurrently [21, 19, 71, 76, 87, 88]. A third type of blind deconvolution algorithms, known as hybrid blind deconvolution, estimates partial information about the PSF which is subsequently used to estimate an optimal inverse filter [41]. Blind deconvolution algorithms can also be classified as parametric [87, 88, 41] or non-parametric [89, 76] based on whether a mathematical model of the PSF is assumed or not. The parametric methods are characterized by estimation of fewer variables and thus are computationally efficient.

Contribution 2: In this thesis, we will propose an efficient numerical solution for the case of hybrid blind deconvolution which is a parametric inverse filtering approach and was originally proposed in [41]. The method in [41] uses a computationally involved smooth optimization technique for the estimation of the inverse filter. On the other hand, the approach that will be proposed in this thesis uses BTS to derive a computationally efficient solution to the problem.

2.1.3 Fast Reconstruction of Diffusion MRI Data

Central to magnetic resonance imaging (MRI) is the notion of contrast, which is typically defined by the biochemical composition of interrogated tissue as well as by the morphology of its associated parenchyma. Prevalent in MRI practice are the contrasts determined by the T_1/T_2 relaxation times and proton density (PD). Despite their exceptional importance to clinical diagnosis, none of the above contrast mechanisms has demonstrated effectiveness in delineating the morphological structure of the white matter. It is only with the advent of dMRI that scientists have been able to perform quantitative measurements of the diffusivity of white matter, based on which its structural delineation has become possible [107, 108, 109, 110, 111, 112, 113]. The dMRI data is collected by repetitive acquisition of MR responses from the same volume for a number of diffusion-encoding gradients, thus measuring the apparent diffusivity along each direction.

Diffusion tensor imaging (DTI) is by far the most widespread type of dMRI used in research and clinical applications [108, 109, 107, 114, 111, 112, 115]. The key assumption employed by DTI is that, at each spatial location, the ensemble average diffusion propagator

can be modelled by a unimodal Gaussian distribution. In other words, it assumes that each voxel can support only one diffusion flow. Unfortunately, the above assumption is known to be violated at the sites where neural fibres cross, touch upon each other or diverge [116, 117, 118, 119].

The fibre crossing problem in DTI has prompted efforts to develop dMRI methodologies which are capable of detecting multiple diffusion flows (or, equivalently, neural fibre tracts) within a given voxel. One of such techniques is high angular resolution diffusion imaging (HARDI)[116, 120, 121, 122, 118, 123, 124, 119, 125], which is capable of capturing multi-modal diffusion patterns. One way of obtaining HARDI data is to sample the whole q-space 3-D Cartesian grid to estimate the ensemble average diffusion propagator. Such a scheme is referred to as diffusion spectrum imaging (DSI) [122, 116, 121] and is often not preferred due to the requirement of large number of samples and large pulsed field gradients [123]. As an alternative, one can sample a spherical shell in the q-space, resulting in Q-ball imaging [123]. However, the required number of samples for this technique, although less than DSI, is still too large to be deemed practical in clinical settings. In particular, the requirement of a large number of samples translates to relatively long acquisition times which greatly impair the practical value of this important imaging modality.

The problem of long acquisition time in HARDI can be overcome by using the theory of compressed sensing (CS) (aka compressive sensing or compressive sampling). The recently developed theory of CS provides a fundamentally new approach of reconstruction of signals from less number of samples than what has been required by the classical sampling theory. The theory has been developed in the works reported in [23, 24, 25, 91, 92, 93, 94] although the fundamental concepts of sparse recovery can be traced back to the work in [95, 96]. The attractive features of compressed sensing have motivated researchers to apply this technique to diverse fields. For example, it has been used to the theory of medical imaging [97, 98, 99, 100], computational biology [101, 102], geophysical data analysis [103, 104], hyperspectral imaging [105], astronomy [106] and several other fields.

The ideas of CS have already paved their way into the field of diffusion imaging [126, 127, 128, 129, 130, 131]. In particular, it has been used to reduce the number of diffusion encoded images required to reconstruct the signal. For example, the work in [129, 130] applies the theory of CS to DSI. On the other hand, the work in [126, 127, 128] applies the theory of CS to HARDI signals sampled on the spherical shell. For example, the approach in [127, 128] uses an exponential mixture model with a sparsity constraint to reconstruct the signal at each voxel. The approach in [126] also performs voxel-wise reconstruction of the signal of interest using CS. But, in contrary to [127, 128], it uses the spherical ridgelet basis [132] to represent the HARDI signal sparsely.

Contribution 3: In this thesis, we will present an approach to perform compressed sensing based reconstruction of HARDI signals sampled on unit sphere from subcritical samples which is important to address the problem of long acquisition time. In contrary to [126, 127, 128], it uses a novel formulation, in which the sparsity constraints enforced in the diffusion domain are augmented by regularity constraints enforced in the spatial domain. The resulting reconstruction problem has the format of a convex minimization problem, which is solved using the technique of BTS. As will be shown in the thesis, the proposed algorithm results in a particularly advantageous computational structure which allows the solution to be computed via a sequence of simple and easily parallelizable steps.

2.2 Splitting Methods for Solving Optimization Problems

Solving large scale optimization problems has been a central topic for researchers for several decades [34, 37]. In this respect, several old methods and their modifications have been exploited in recent years for solving large scale problems as well as making the methods parallelizable [133, 134, 32]. The method of obtaining the solutions is characterized by superseding the original optimization problem by a number of simpler ones using the notion of splitting. There are several types of splitting that can be used in a wide variety of practical scenarios [34, 29, 133, 134, 135].

In this thesis, we are interested in two specific types of splitting. The first one is the forward-backward splitting [34]. The method has also been derived recently from different theoretical considerations in a number of works [29, 133, 4, 136, 135]. Thus for example, the forward backward-splitting method can be derived as an expectation maximization algorithm [4], as a majorization-minimization method [136] or as an iterative shrinkage approach [29]. The second type of splitting is the method of split Bregman which has been motivated by the Bregman distance proposed in [137]. The original idea has been extended to solve large scale problems in several works reported in [138, 33, 31]. In parallel, the method of ADMM has also been used to solve the same type of problems [32, 22]. In fact, it has been shown in the work of [31] that the two approaches are identical and one can be derived from the other. The splitting method simplifies a computationally complex problem with multiple regularizers leading to an attractive, computationally efficient solution structure. Consequently, the approaches have been used in several applications to solve large scale optimization problems in the context of deconvolution and compressed sensing [22, 32, 33]. In particular, the work in [32, 33] addresses problems where the images

are contaminated with Gaussian noise whereas the work in [22] extends the approach to the case of images contaminated by Poisson noise.

Contribution 4: In this thesis, we will show the application of Split Bregman/ADMM approach to simplify complex optimization problems for the case of compressed sensing and image deconvolution. Specifically, in each case, we will address problems where the reconstruction formulations are characterized by cross-domain constraints leading to non-trivial optimization problems. In such a scenario, the splitting method results in simpler subproblems whose fast/efficient solution methods are known.

Chapter 3

Technical Preliminaries

In this chapter, we provide several technical concepts that will be useful in the presentation of our research.

3.1 Convex Optimization

As discussed in Chapter 1, many of the estimation methods require solving an *optimization problem*. Such an optimization problem can be derived from an intuitive understanding of the problem as well as using a formal Bayesian framework. An optimization problem has a general form

$$\begin{aligned} \min_{x} \;\; & f_0(x) \\ \text{s.t.} \;\; & f_i(x) \leq b_i, \qquad i = 1, 2, \ldots, m \end{aligned} \tag{3.1}$$

In the above formulation, the vector $x = (x_1, \ldots, x_N)$ is referred to as the *optimization variable* of the problem, the function $f_0 : \mathbb{R}^N \mapsto \mathbb{R}$ is the objective or cost function, whereas the functions $f_i : \mathbb{R}^N \mapsto \mathbb{R}$, $i = 1, \ldots, m$, are called the (inequality) constraint functions. The constants $b_1, \ldots, b_m$ in (3.1) are the limits, or bounds, for the constraints. A vector $x^* \in \mathbb{R}^N$ is called an *optimal* or a *solution* to the problem in (3.1) if it results in the smallest value of the objective function among all the vectors that satisfies the constraints.

A specific class of optimization problem is the *convex optimization* problem [139, 28]. In this case, the objective function as well as the constraint functions are *convex*. In order

to formally define a convex function, let us first define the notion of a *convex set*. A set $\mathcal{C}$ is defined to be convex if

$$\theta x + (1-\theta)y \in \mathcal{C}, \quad \forall x, y \in \mathcal{C} \text{ and any } \theta \text{ with } 0 \leq \theta \leq 1. \tag{3.2}$$

Another definition that is useful in this context is the notion of domain of an extended real valued function $f : \mathbb{R}^N \mapsto (-\infty, \infty]$ which is defined as

$$\text{dom}(f) = \{x \in \mathbb{R}^N | f(x) < \infty\} \tag{3.3}$$

Subsequently, a function $f : \mathbb{R}^N \mapsto \mathbb{R}$ is convex if $\text{dom}(f)$ is a convex set and if

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y), \quad \forall x, y \in \text{dom}(f),\ 0 \leq \theta \leq 1 \tag{3.4}$$

Geometrically, this implies that the line segment between $(x, f(x))$ and $(y, f(y))$, which is the chord from x to y lies above the graph of f. Moreover, a function f is called strictly convex if (3.4) holds with strict inequality.

In convex analysis, a number of different important definitions will be used in this thesis and accordingly they are introduced here. A central concept in convex analysis is that of subgradient ∂f, which, at $x \in$dom(f), is defined as

$$\partial f(x) = \{u \in \mathbb{R}^N \mid \forall y \in \mathbb{R}^N, f(y) \geq f(x) + \langle y - x, u \rangle\} \tag{3.5}$$

and $\langle \cdot, \cdot \rangle$ denote the inner product over $\mathbb{R}^N$ defined as

$$\langle x, y \rangle = \sum_{i=1}^{N} x_i y_i \tag{3.6}$$

If the function f is Gâteaux differentiable at the point $x \in \text{dom}(f)$ with gradient $\nabla f(x)$, then $\partial f(x) = \{\nabla f(x)\}$ is a singleton set.

A function f is said to be proper if $\text{dom}(f) \neq \emptyset$, i.e. f is not identically equal to ∞. Finally, a function f is called lower semicontinuous at a point $x \in \mathbb{R}^N$ if

$$f(x) \leq \liminf_{k \to \infty} f(x_k) \tag{3.7}$$

for every sequence $\{x_k\} \subset \mathbb{R}^N$ that converges to x.

The advantage of a strictly convex optimization problem lies in the fact that it has a unique optimal point. If x^* is the unique optimal point, then

$$0 \in \partial f(x^*) \tag{3.8}$$

If the function is differentiable at x^*, then

$$\nabla f(x^*) = 0. \tag{3.9}$$

The field of convex optimization has been highly advanced and consequently in many cases, many non-convex problems are also treated using the tools of convex optimization. In such cases, one trades off the possibility of not attaining the optimal solution with the simplicity of the available tools.

A very simple numerical method to obtain the optimal point is the steepest descent algorithm and we are going to discuss this to show the connection with the proximity operators introduced in the next section. Let $f : \mathbb{R}^N \mapsto \mathbb{R}$ be a continuously differentiable convex function. The steepest descent algorithm finds the optimal solution iteratively according to

$$x^{k+1} = x^k - \gamma \nabla f(x^k), \tag{3.10}$$

where γ is a step size and $\nabla f(x)$ is the gradient evaluated at the point x. It should be noted that the gradient $\nabla f(x^k)$ in (3.10) has been computed at the current iterate x^k and such a scheme is called an explicit scheme. There can be a corresponding implicit scheme where the gradient is computed based on the next iterate, i.e. at the point x^{k+1} and this will be of interest in the next section.

3.2 Proximity Operations and Related Solutions

In order to introduce the notion of proximity operators, let us first discuss the concept of projection onto convex sets [140]. Let $\mathcal{C}$ be a non-empty closed convex set of the real Hilbert space $\mathbb{R}^N$ with the Euclidean norm $\|\cdot\|$. It should be noted that although the content is presented using the Hilbert space $\mathbb{R}^N$, a general Hilbert space with its associated inner product can also be used. For every $y \in \mathbb{R}^N$, there exists a unique point $P_{\mathcal{C}}(y) \in \mathcal{C}$, called the projection of y onto $\mathcal{C}$, which satisfies the best approximation property

$$\|y - P_{\mathcal{C}}\, y\| \leq \|y - z\|, \quad \forall z \in \mathcal{C} \tag{3.11}$$

It should be noted that $P_{\mathcal{C}}\, y$ can also be written as the unique solution to the variational problem

$$\min_{x \in \mathbb{R}^N} \imath_{\mathcal{C}}(x) + \frac{1}{2}\|x - y\|^2 \tag{3.12}$$

where $\imath$ is the indicator function of $\mathcal{C}$ defined as

$$\imath_{\mathcal{C}}(x) = \begin{cases} 0, & \text{if } x \in \mathcal{C} \\ +\infty, & \text{otherwise} \end{cases} \tag{3.13}$$

The indicator function in (3.12) has been subsequently replaced by more general functions in [141] to introduce the notion of proximity operators. To put things formally, let $\Gamma_0(\mathbb{R}^N)$ be the set of convex, lower semicontinuous, proper functions defined from $\mathbb{R}^N$ to $\mathbb{R} \cup \{+\infty\}$ and $g \in \Gamma_0(\mathbb{R}^N)$. Then, for every $y \in \mathbb{R}^N$, the proximity operator $\text{prox}_g : \mathbb{R}^N \mapsto \mathbb{R}^N$ is defined as

$$\text{prox}_g(y) = \arg\min_{x \in \mathbb{R}^N} g(x) + \frac{1}{2}\|x - y\|^2 \tag{3.14}$$

Specifically, it has been shown in [141] that the problem in (3.14) admits a unique solution.

An interesting property of the proximity operator is that [34, 134]

$$\forall y, p \in \mathbb{R}^N, \quad p = \text{prox}_{\gamma g}(y) \iff y - p \in \gamma\, \partial g(p) \tag{3.15}$$

which reduces to

$$\forall y, p \in \mathbb{R}^N, \quad p = \text{prox}_{\gamma g}(y) \iff y - p = \gamma\, \nabla g(p) \tag{3.16}$$

if g is differentiable. This can be rearranged as

$$p = y - \gamma\, \nabla g(p) \tag{3.17}$$

With reference to the previous section, one can identify that this is the backward (implicit) method of minimizing a convex function. The forward and backward steps of minimizing a convex function are coupled in the iterative shrinkage algorithm which has been discussed in the next section.

The notion of proximity operator is very important in this thesis and we are going to derive the proximity operators for the following cases.

3.2.1 Proximity Operator for ℓ_2 Norm

Substituting

$$g(x) = \frac{\gamma}{2}\|\Gamma x - z\|^2, \tag{3.18}$$

where $\gamma > 0$, $z \in \mathbb{R}^N$ and $\Gamma \in \mathbb{R}^{N \times N}$, in (3.14), we have the following minimization problem

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - y\|^2 + \frac{\gamma}{2}\|\Gamma x - z\|^2 \tag{3.19}$$

We can also consider a more general version of the problem given by

$$\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Ax - y\|^2 + \frac{\gamma}{2}\|\Gamma x - z\|^2 \tag{3.20}$$

where $A \in \mathbb{R}^{N \times N}$. To minimize the expression in (3.20), we compute the gradient with respect to x and set the result equal to zero as given by,

$$A^T(Ax - y) + \gamma\Gamma^T(\Gamma x - z) = 0 \tag{3.21}$$

Consequently, after rearrangement, the solution is given by

$$x = (A^T A + \Gamma^T \Gamma)^{-1}(A^T y + \Gamma^T z) \tag{3.22}$$

In the case when $z = 0$, the regularizer $\frac{1}{2}\|\Gamma x\|^2$ is called the Tikhonov's regularizer. In addition, if $\Gamma = I$, the identity matrix, a solution with smaller norm is preferred. This regularizer is often used when the matrix A is ill-conditioned and hence leads to a non-unique or unstable solution when used without a regularizer. This is the underlying philosophy of the Wiener filter [1].

3.2.2 Proximity Operator for ℓ_1 Norm

In this case, the optimization problem is given by

$$\arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|x - y\|_2^2 + \gamma\|x\|_1 \tag{3.23}$$

where $\gamma > 0$, and

$$\|x\|_1 = \sum_{i=1}^{N} |x_i| \tag{3.24}$$

with x_i being the ith element of the vector x. The problem in (3.23) can be written element-wise as

$$\min_{x_i} \frac{1}{2}(x_i - y_i)^2 + \gamma|x_i| \tag{3.25}$$

To minimize (3.25), we find the derivative of the expression and set it equal to zero. Hence, for $x_i \neq 0$

$$x_i + \gamma \operatorname{sign}(x_i) = y_i \tag{3.26}$$

Since $|x_i|$ is not differentiable at $x_i = 0$, the derivative is replaced by subdifferential. In particular, the subdifferential of $|x_i|$ at the point $x_i = 0$ is

$$\partial |x_i| \Big|_{x_i=0} = [-1, +1]. \tag{3.27}$$

Hence for $x_i = 0$

$$x_i - y_i \in [-\gamma, +\gamma] \implies y_i \in [-\gamma, \gamma] \quad \text{for } x_i = 0 \tag{3.28}$$

The plot of y_i w.r.t x_i is shown in Fig. 3.1(a). The simplest way to express x_i in terms

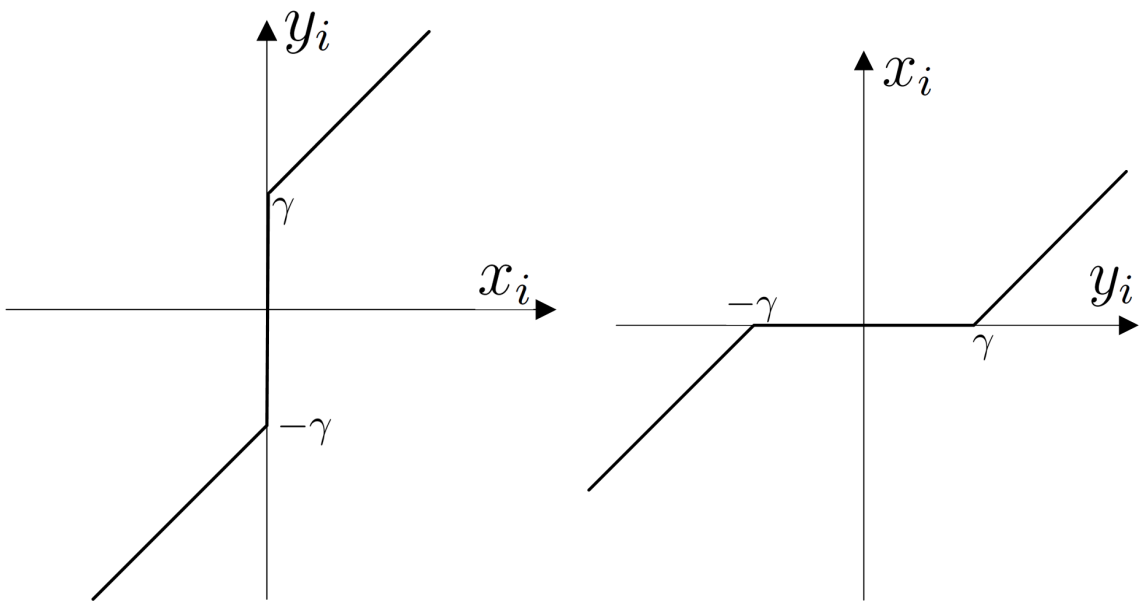


Figure 3.1: (a) Combined plot of (3.26) and (3.28) (b) soft thresholding function

of y_i is to change the coordinate system and the plot in Fig. 3.1(a). The resulting plot is shown in Fig. 3.1(b). The solution of (3.25) is known as *soft thresholding* and formally it can be written as

$$x_i = \operatorname{soft}_\gamma(y_i) = \frac{\max(|y_i| - \gamma, 0)}{\max(|y_i| - \gamma, 0) + \gamma} y_i \tag{3.29}$$

Note that, although the above expression has been derived assuming x_i and y_i are real, the same derivation as well as the result in (3.29) holds for complex variables as well [142].

3.2.3 Proximity Operator for TV Seminorm

In this subsection, we are going to consider the case when the regularizer is given by the *total variation* (TV) seminorm. For an image modelled as a differentiable function $f \in \mathcal{C}^1(\Omega)$ on a domain $\Omega \subset \mathbb{R}^N (N \geq 1)$, the total variation seminorm is defined as

$$\|f\|_{\mathrm{TV}} = \int_{\Omega} |\nabla f(x)| dx \tag{3.30}$$

where ∇ denotes the gradient operator and $|\cdot|$ denotes the Euclidean norm. The above definition requires the function to be differentiable which limits the class of function where the definition can be used. The definition can be extended to locally integrable functions using a dual formulation. Formally, for $f \in \mathbb{L}^1_{loc}(\Omega)$ (the set of all functions whose Lebesgue integral is finite on all compact subsets of Ω),

$$\|f\|_{\mathrm{TV}} = \sup \left\{ \int_{\Omega} f(x) \, \mathrm{div}\xi(x) dx : \xi \in C^1_c(\Omega, \mathbb{R}^N), |\xi(x)| \leq 1, \forall x \in \Omega \right\}, \tag{3.31}$$

where C^1_c is the space of smooth functions with compact support. The definition of the total variation seminorm is related to the concept of bounded variation functions. In particular, a function is said to be of bounded variation if its total variation is finite.

Unfortunately, the proximity operator using the total variation seminorm does not admit a closed form solution. Accordingly, one needs to resort to iterative approaches for solving the optimization problem. Fortunately, there are fast solvers which can be readily used to find the proximity operator [143, 144, 33]. In particular, we use the approach of [143] in this thesis to compute the proximity operator corresponding to TV seminorm. It should be noted that the approach in [143] uses a discretization of the dual definition of total variation defined in (3.31).

3.3 Iterative Shrinkage

The iterative shrinkage algorithm considers optimization problems of the form

$$\arg\min_x \{F(x) \equiv f(x) + g(x) : x \in \mathbb{R}^n\} \tag{3.32}$$

where

1. $g : \mathbb{R}^n \mapsto \mathbb{R}$ is a continuous convex function, and
2. $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a continuous convex function which is continuously differentiable with Lipschitz continuous gradient $L(f) > 0$, i.e.

$$\|\nabla f(x) - \nabla f(y)\| \leq L(f)\|x - y\|, \quad \forall x, y \in \mathbb{R}^n \tag{3.33}$$

where $\|\cdot\|$ denotes the standard Euclidean norm.

The algorithm has been derived using different frameworks [136, 29, 135, 4, 37, 133]. The solution of this problem is obtained iteratively. Let x^k be the solution at some arbitrary time step k. Then the next iterate x^{k+1} is obtained as

$$x^{k+1} = \arg\min_x Q_L(x, x^k) \tag{3.34}$$

where

$$Q_L(x, x^k) = f(x^k) + \langle x - x^k, \nabla f(x^k)\rangle + \frac{L}{2}\|x - x^k\|^2 + g(x) \tag{3.35}$$

has the following properties

1. $Q_L(x^k, x^k) = f(x^k) + g(x^k) = F(x^k)$, i.e. the value of $Q_L(x, x^k)$, evaluated at $x = x^k$ is equal to the value of $F(x)$ evaluated at $x = x^k$.
2. $Q_L(x, x^k) \geq F(x)$, i.e. $Q_L(x, x^k)$ upper bounds (majorizes) $F(x)$ touching it for $x = x^k$.

Ignoring the terms which does not depend on x, the minimization process can also be written as

$$x^{k+1} = p_L(x^k) \tag{3.36}$$

where

$$p_L(y) = \arg\min_x \left\{ \frac{L}{2} \left\| x - \left(y - \frac{1}{L}\nabla f(y) \right) \right\|^2 + g(x) \right\} \tag{3.37}$$

It should be noted that $p_L(x_k)$ can be decomposed as:

1. A forward Euler step of length $1/L$ in the direction of the negative gradient of the data fidelity term as given by

$$x^{k+1/2} = x^k - \frac{1}{L}\nabla f(x^k) \tag{3.38}$$

2. A backward (or implicit) or proximal operation step with reference to the regularizer $g(x)$ as given by

$$x^{k+1} = \text{prox}_{\frac{1}{L}g}(x^{k+1/2}) \tag{3.39}$$

The name, forward-backward splitting, has been coined due to the peculiarity of the method's computational organization.

There are situations when the Lipschitz constant $L(f)$ is not always known or computable. In such a case, a backtracking method [135] can be used as shown in Algorithm 1.

Algorithm 1 Iterative shrinkage with backtracking

Input: Take $L^0 > 0$: some $\eta > 1$ and $x^0 \in \mathbb{R}^n$

Step k ($k \geq 1$): Find the smallest non-negative integer c^k, such that with $\bar{L} = \eta^{c^k} L^{k-1}$

$$F(p_{\bar{L}}(x^{k-1})) \leq Q_{\bar{L}}(p_{\bar{L}}(x^{k-1}), x^{k-1}) \tag{3.40}$$

Set $L^k = \eta^{c^k} L^{k-1}$ and compute

$$x^k = p_{L^k}(x^{k-1}) \tag{3.41}$$

The iterative shrinkage algorithm is very simple and hence is attractive for solving large-scale problems. However, the method is known to converge only linearly in the vicinity of an optimal solution [135]. An improvement has been proposed in [135] which has been referred as *fast iterative shrinkage-thresholding algorithm* (FISTA). The algorithm has been shown to converge quadratically [135]. The steps of the algorithm have been shown in Algorithm 2.

3.4 Neighbourhood Filtering

In this section, we consider different types of spatial-domain filtering because of their importance in this thesis. The filtering operation in the spatial domain is often referred to as neighbourhood filtering. The name roots from the fact that the intensity of a pixel is modified based on the intensities of the pixels in the neighbourhood of the original pixel. In order to specify the important neighbourhood filters, let us first introduce the following notations. Let $\Omega \subset \mathbb{Z}^2$ denotes the domain of definition of the images under consideration. Also, let f_z and g_z denote the intensities of the original image f and the observed noisy

Algorithm 2 Fast iterative shrinkage with backtracking

Input: Take $L^0 > 0$: some $\eta > 1$ and $x^0 \in \mathbb{R}^n$. Set $y^1 = x^0$, $t^1 = 1$.
Step k ($k \geq 1$): Find the smallest non-negative integer c^k, such that with $\bar{L} = \eta^{c^k} L^{k-1}$

$$F(p_{\bar{L}}(y^k)) \leq Q_{\bar{L}}(p_{\bar{L}}(y^k), y^k) \tag{3.42}$$

Set $L^k = \eta^{c^k} L^{k-1}$ and compute

$$x^k = p_{L^k}(y^k) \tag{3.43}$$

$$t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2} \tag{3.44}$$

$$y^{k+1} = x^k + \frac{t^k - 1}{t^{k+1}}(x^k - x^{k-1}) \tag{3.45}$$

image g respectively, at the spatial location $z \in \Omega$. Then the neighbourhood filtering operation estimates $\hat{f}_z$ at $z \in \Omega$ from g according to

$$\hat{f}_z = \frac{1}{C_z} \sum_{\xi} w_{z,\xi} g_\xi \tag{3.46}$$

where $C_z = \sum_\xi w_{z,\xi}$. Depending on the choice of $w_{z,\xi}$, the following important types of filters can be obtained.

- **Linear Space Invariant Filter:** In the case of linear filter, $w_{z,\xi}$ is a function of the orientation and distance of the pixel at location ξ with respect to that at z. As a specific example, we can consider the Gaussian filter where the weights are given by

$$w_{z,\xi} = \exp\left\{-\frac{\|z - \xi\|^2}{2\rho_s^2}\right\}, \tag{3.47}$$

  and ρ_s controls the decay of the filter coefficients with increasing spatial distance.

- **Range Filter:** In the case of range filter, the filter weights are given by

$$w_{z,\xi} = \exp\left\{-\frac{(g_z - g_\xi)^2}{2\rho_r^2}\right\} \tag{3.48}$$

  where the parameter ρ_r in (3.48) controls the amount of filtering. High values of ρ_r tend to result in an overly smoothed image, whereas a smaller value results in a

milder filtering effect. Hence, this parameter ρ_r should be set to maintain a balance between the removal of noise and preservation of details of the image. As a general rule, the optimal value of ρ_r can be estimated based on the level of the contaminating noise [38].

- **Bilateral Filter:** In the case of bilateral filter, the filter weights are given by

$$w_{z,\xi} = \exp\left\{-\frac{\|z-\xi\|^2}{2\rho_s^2}\right\}\exp\left\{-\frac{(g_z-g_\xi)^2}{2\rho_r^2}\right\} \tag{3.49}$$

  where ρ_s and ρ_r control the decay of the filter coefficients with increasing spatial distance and difference in intensities respectively.

- **Non-Local Means Filter:** For non-local means filter, the weights are given by

$$w_{z,\xi} = \exp\left(-\frac{1}{h}\sum_{k\in\Gamma}\beta_k|g_{z-k}-g_{\xi-k}|^2\right) \tag{3.50}$$

  with Γ defined in such a way that $\{g_{z-k}\}_{k\in\Gamma}$ (resp. $\{g_{\xi-k}\}_{k\in\Gamma}$) represents a symmetric (rectangular) neighbourhood of the source intensity g_z (resp. the target intensity g_ξ). The parameters $\{\beta_k\}_{k\in\Gamma}$ in (3.50) are intended to weight the domain of summation and they should be chosen such that $\sum_{k\in\Gamma}\beta_k = 1$, while $h > 0$ as before controls the overall amount of smoothing imposed by the filter.

3.5 Compressed Sensing

The conventional techniques of sampling an analog signal are guided by the famous Shannon/Nyquist theorem which guarantees the signal reconstruction if the sampling rate is at least twice that of the maximum frequency present in the signal. This assumes that the signal is bandlimited. For the signals, which are not naturally bandlimited, the sampling rate is determined by the required resolution and in this case, the signal is first passed through a low pass filter to suppress the high frequencies and then the Shannon/Nyquist theorem is used. However, recent research has shown that a signal can be recovered from far fewer numbers of samples than required by the classical sampling theory. The underlying theory is known as compressed sensing (CS) [23, 24, 25, 91, 92, 93, 94].

For the sake of simplicity, let us present the theory of CS for finite length discrete signals. In particular, let $f \in \mathbb{R}^N$ is the quantity of interest (note that a 2-dimensional

image can be arranged in a lexicographic order resulting in an 1-dimensional vector). The measurement model is assumed to be linear and the measurements are given by

$$y_i = \langle f, \phi_i \rangle, \quad i = 1, 2, \ldots, M \tag{3.51}$$

where ϕ_i are predefined sampling functions. Thus, for example, the function ϕ_i can be dirac delta (conventional sampling) or complex exponentials (as in MRI). We are interested in an undersampled situation where $M \ll N$. The question is whether the signal f can be reconstructed from M acquisitions and the theory of CS answers the question affirmatively provided some conditions are satisfied.

The theory of CS depends on two key concepts: "sparsity" and "incoherence", each of which is discussed below.

Sparsity: The concept of sparsity says that the degree of freedom of a discrete signal is much smaller than the length of the signal. This relies on the fact that most of the natural signals are sparse or compressible; in other words, they admit a sparse representation in the domain of some suitably designed linear transform Ψ. Let us assume that Ψ is the basis matrix of orthonormal vectors $\{\psi_i\}_{i=1}^N$ as given by $\Psi = [\psi_1, \psi_2, \cdots, \psi_n]$ and any signal $f \in \mathbb{R}^N$ can be represented as a linear combination of the basis functions according to

$$f = \sum_{i=1}^{N} x_i \psi_i \quad \text{or equivalently } f = \Psi x \tag{3.52}$$

Here $x \in \mathbb{R}^N$ is a vector of representation coefficients and since the basis is orthonormal $x_i = \langle f, \psi_i \rangle$. The signal f is *K-sparse* if it is a linear combination of only K basis vectors; i.e. only K of the above N coefficients are non-zero and the remaining $N - K$ coefficients are zero. Denoting x_K to be the vector of coefficients with all but K largest coefficients set to zero and $f_K = \Psi x_K$, the signal f is said to be *compressible* if the error $\|f - f_K\|_{\ell_2} = \|x - x_K\|_{\ell_2}$ is small.

The sparsity or compressibility of the signal is the key-factor for compressing a signal. In particular, if the signal has K large coefficients, one can store the values and locations of the K coefficients and discard the remaining $N - K$ small coefficients. However this method has the following disadvantages;

1. One has to acquire all the N samples of the signal and N can be very large.
2. The set of all the N coefficients has to be computed although the $N - K$ coefficients will be discarded.

3. The indices of the K large coefficients are dependent on the signal. In that case, one has to store the indices of the coefficients thus introducing overhead.

The theory of CS, on the other hand, allows acquiring M ($M \ll N$) in a non-adaptive way.

Incoherence: Let us denote the full $N \times N$ sensing matrix as Φ which is composed of the acquisition vectors ϕ_k. Consequently, let us define coherence between Φ and Ψ as

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{1 \leq k,j \leq N} |\langle \phi_k, \psi_j \rangle| \tag{3.53}$$

The theory of CS requires a sensing matrix which has a low coherence with the representation matrix Ψ. This implies that the signal having sparse representation in the Ψ domain has a dense representation in the Φ domain or equivalently, the sampling waveforms ($\{\phi_i\}_{i=1}^N$) have a dense representation in Ψ.

In the CS setting, one directly acquires M ($M \ll N$) random samples by projecting on M vectors $\{\phi_i\}_{i=1}^M$ as given by (3.51). Subsequently, one can perform the reconstruction according to

$$\min_{\tilde{x} \in \mathbb{R}^N} \|\tilde{x}\|_{\ell_1}, \quad \text{subject to } y_i = \langle \phi_i, \Psi \tilde{x} \rangle, \quad i = 1, \ldots, M \tag{3.54}$$

It has been shown in [145] that for an K-sparse signal, if

$$M \geq C\, \mu^2(\Phi, \Psi)\, K\, \log(N) \tag{3.55}$$

then the solution of (3.54) is exact with overwhelming probability. From (3.55), one can observe that smaller the coherence μ, the fewer the samples required. If $\mu(\Phi, \Psi) = 1$, then on the order of $K \log(N)$ samples are required for accurate reconstruction as opposed to N.

There are stronger version of the above result [24]. To state that, let us first denote

$$\Theta = R\Phi\Psi, \tag{3.56}$$

where Φ and Ψ are the acquisition and representation bases respectively and R is an $M \times N$ matrix which represents the subsampling operation. Next, let us state the following property [146].

Restricted Isometry Property (RIP) For each integer $K = 1, 2, \ldots$, define the isometry constant δ_K of a matrix Θ as the smallest number such that

$$(1 - \delta_K)\|x\|_{\ell_2}^2 \leq \|\Theta x\|_{\ell_2}^2 \leq (1 + \delta_K)\|x\|_{\ell_s}^2 \tag{3.57}$$

holds for all K-sparse vectors x. Subsequently, Θ is said to follow the RIP of order k if δ_k is not very close to 1. This implies that all subsets of k columns taken from Θ are nearly orthogonal and it preserves the Euclidean length of k-sparse signals.

Subsequently, we state the following theorem [24].
Theorem: Assume that $\delta_{2K} < \sqrt{2} - 1$. Then the solution x^* to (3.54) obeys

$$\begin{aligned} \|x^* - x\|_{\ell_1} &\leq C_0 \|x - x_K\|_{\ell_1} \\ \|x^* - x\|_{\ell_2} &\leq C_0 \|x - x_K\|_{\ell_1} / \sqrt{K} \end{aligned} \tag{3.58}$$

for some constant C_0 where x_K is the vector x with all but the largest K components set to zero. The conclusion of the above theorem is as follows. If x is K-sparse, then $x = x_K$ and the recovery is exact. However, if x is not K-sparse, then (3.58) guarantees that the quality of the recovered signal is as good as if one knows ahead of time the position of the K largest coefficients and perform the reconstruction based on those coefficients.

To present the CS theory for noisy data, let us consider the following measurement model

$$y = \Theta x + \eta \tag{3.59}$$

where $\eta \in \mathbb{R}^M$ is the contaminating noise. In such a case, the CS theory proposes reconstruction of x by solving the optimization problem

$$\min_{\tilde{x} \in \mathbb{R}^N} \|\tilde{x}\|_{\ell_1}, \quad \text{subject to } \|\Theta \tilde{x} - y\|_{\ell_2} \leq \epsilon \tag{3.60}$$

where $\epsilon > 0$ depends on the level of the contaminating noise. Subsequently, we have the theorem from [24].
Theorem: Assume that $\delta_{2K} < \sqrt{2} - 1$. Then the solution x^* to (3.60) obeys

$$\|x^* - x\|_{\ell_2} \leq C_0 \|x - x_K\|_{\ell_1} / \sqrt{K} + C_1 \epsilon \tag{3.61}$$

for some constants C_0 and C_1. Hence the reconstruction error is bounded by the sum of two terms; the term corresponding to the noise-free case and a term proportional to the noise level.

Chapter 4

Compressed Sensing with Cross-Domain Constraints using BTS

In this chapter, we are going to address a specific type of inverse problems which is very important in the context of image reconstruction. In particular, it addresses the issue of incorporating *multiple composite constraints* within the reconstruction framework, where each of the constraints can arise from a priori information defined over different domains of definition of the image. Subsequently, we are going to use this framework to address some specific applications in the case of multidimensional vector valued images. Complete description of such type of images is associated with acquisition of large amount of data, which is often not feasible due to practical constraints. Accordingly, in order to reduce the amount of acquired data in such cases, the theory of compressed sensing has been proposed and the computational complexity of solving the optimization problem has been simplified using the principles of BTS. As an example of this proposed reconstruction framework, the problem of long acquisition time inherent in diffusion MRI has been addressed.

4.1 Model Specification and Motivation

Let us consider the following mapping

$$s : \Omega_1 \mapsto \mathcal{H}(\Omega_2) : \ \mathbf{r} \mapsto s(\mathbf{r}) \tag{4.1}$$

where Ω_1 is the domain of definition of the function s whereas $\mathcal{H}$ is a Hilbert space defined over Ω_2 with its associated definition of the inner product. In the definition in (4.1), the

function s maps each point of $\mathbf{r} \in \Omega_1$ to a function $s(\mathbf{r}) \in \mathcal{H}(\Omega_2)$. When combined together, the continuum of vectors $s(\mathbf{r})$ can be regarded as a vector field in which case $s(\mathbf{r})$ has a natural interpretation of the value of s corresponding to position $\mathbf{r}$.

It should be noted that the signal can also be interpreted as a function defined over $\Omega_1 \times \Omega_2$ and the map can be represented by

$$s : \Omega_1 \times \Omega_2 \mapsto \mathbb{R} : \ (\mathbf{r}, \mathbf{u}) \mapsto s(\mathbf{u} \mid \mathbf{r}), \ \mathbf{r} \in \Omega_1, \ \mathbf{u} \in \Omega_2 \tag{4.2}$$

Before stating the problem formally, let us first provide some motivating examples which fall under the category of function definition shown in (4.1).

4.1.1 Color Images

The most common example that fits the model in (4.1) is that of colour images. In this case, the image is defined over a two dimensional plane, i.e. $\Omega_1 \subset \mathbb{R}^2$. At each spatial location, the image is defined by three intensity values, namely red, green and blue (often referred to as RGB color scheme). Hence, Ω_2 can be defined as $\Omega_2 = \{1, 2, 3\}$ and the range of the mapping in (4.1), i.e. the Hilbert space $\mathcal{H}(\Omega_2)$ is the set of real-valued triplets.

4.1.2 Hyperspectral Imaging

A more advanced form of RGB colour base image is hyperspectral (HS) imaging which refers to imaging the electromagnetic wave properties of a scene or an object. As opposed to a regular camera which acquires an image with low spectral resolution (RGB colour base), the HS imaging is characterized by a much wider spectral range (ultra-violet to deep IR) and much higher spectral resolution (from one hundred to two thousand spectral bands). The HS imaging is a very powerful technique to identify and quantify distinct material substances from observed spectral data. Consequently, it has a wide range of applications such as terrain classification, mineral detection and exploration, pharmaceutical counterfeiting, environmental monitoring and military surveillance [147, 148, 149, 150].

The HS imaging can also be described by the mapping in (4.1). Similar to the case of color images, $\Omega_1 \subset \mathbb{R}^2$ denotes the domain of definition of the HS image, whereas $\Omega_2 \subset \mathbb{R}^+$ denotes the set of wavelengths which are represented within the acquired images. The Hilbert space $\mathcal{H}$ is the set of all square integrable functions defined on $\mathbb{R}^+$ and formally,

$$\mathcal{H}(\Omega_2) = \mathbb{L}_2(\mathbb{R}^+) \tag{4.3}$$

4.1.3 Diffusion MRI

Another important application which can be described by the mapping in (4.1) is the case of diffusion MRI. Diffusion MRI can delineate the morphological structure of the white matter of the human brain by tracing the diffusion of water molecules within neural tracts located in that region. A very important type of diffusion MRI technique is *high angular resolution diffusion imaging* (HARDI) [151, 123, 120]. Compared to the more traditional *diffusion tensor imaging* (DTI), HARDI is often considered to be more accurate and has the potential of providing more insight to the way the human brain is connected.

In the case of HARDI acquisition, Ω_1 can be considered to be a subset of $\mathbb{R}^3$. At each spatial location $\mathbf{r} \in \Omega_1$, the signal is acquired over a unit sphere which is formally defined as

$$\mathbb{S}^2 = \{\mathbf{v} \in \mathbb{R}^3 \mid \|\mathbf{v}\|_2 = 1\} \tag{4.4}$$

In other words, $\Omega_2 = \mathbb{S}^2$. The Hilbert space $\mathcal{H}$ is assumed to be the space of all square integrable functions defined on the unit sphere, i.e.

$$\mathcal{H}(\Omega_2) = \mathbb{L}_2(\mathbb{S}^2) \tag{4.5}$$

4.1.4 Wireless Sensor Networks

A different example of the mapping in (4.1) is that of the wireless sensor network (WSN) which uses miniature, power efficient and low-cost sensors to monitor different attributes of the regions where they are deployed. In particular, the sensors record its data over time and communicate with the base station using different networking protocols. The WSN is becoming increasingly important and its applications include, yet not limited to, public access multimedia, traffic, security and surveillance, environmental radiation control and many others [152, 153, 154, 155, 156, 157, 158, 159, 160].

In our formal settings as given by (4.1), Ω_1 describes the domain of deployment of the sensor nodes, which in general can be a graph. On the other hand, depending on the application, the set Ω_2 can vary. If we consider the case when the sensors are recording data with respect to time, then $\Omega_2 \subset \mathbb{R}^+$ denotes the time axis. The Hilbert space in this case can be considered to be the set of square integrable function defined on $\mathbb{R}^+$, i.e.

$$\mathcal{H}(\Omega_2) = \mathbb{L}_2(\mathbb{R}^+) \tag{4.6}$$

4.2 Problem Formulation

As discussed in the previous section, the model in (4.1) is common in a number of very important applications. The problem with many of these applications is the fact that one needs to acquire a large amount of data over Ω_2 for each point in Ω_1 and this is often not possible due to practical constraints. Although the problem is not so severe in the case of colour images, it often imposes restrictions in the other scenarios mentioned above. For example, in the case of HS imaging, one needs to acquire the data for a large spectral range which needs a lot of acquisition time and it also leads to relatively large amount of data. Similarly, during HARDI acquisition, the long acquisition time for collecting data over Ω_2 makes its application impractical because of non-compliance of the patient. Moreover, it also leads to degradation in the quality of the data because of motion exhibited by the patient. Likewise, the acquisition of large amount of data is often not feasible in the case of WSN because it suffers from the problem of limited power.

A possible solution to the problem mentioned above is to use the theory of compressed sensing. In general, real life data is highly compressible when expressed in appropriate domain. In other words, they can be expressed as a linear combination of a small number of functions when expressed in a domain of a properly selected transform (often referred to as *sparsifying transform*). Based on the availability of a sparsifying transform defined over $\mathcal{H}(\Omega_2)$, one can use the theory of compressed sensing to reconstruct the data over Ω_2 for each point in Ω_1. It should be noted that one can design sparsifying transform and apply the theory of compressed sensing on the whole data defined over $\Omega_1 \times \Omega_2$. However, in many practical applications, it is often non-trivial to design a sparsifying transform for the whole data. Moreover, in this case, one needs to work with data of a relatively large dimensionality which might be computationally intractable. On the other hand, performing compressed sensing only over Ω_2 allows the algorithm to be parallelizable and hence can be implemented by efficient means. Consequently, we are going to concentrate on the framework of compressed sensing of functions defined over Ω_2.

Before proceeding further, we note that in most of the practical cases, one works with data defined over a discrete domain. Accordingly, it makes sense to assume Ω_1 and Ω_2 to be discrete and in this case, $s(\mathbf{r})$ can be treated as a discrete vector field. Let us represent $s(\mathbf{r})$ as

$$s(\mathbf{r}) = \Psi_{\mathbf{r}}\{c(\mathbf{r})\} \tag{4.7}$$

where $\Psi_{\mathbf{r}}$ is a representation basis for functions defined on $\mathcal{H}(\Omega_2)$ and $c(\mathbf{r})$ are the coefficients corresponding to $s(\mathbf{r})$ in the domain of the transform. It should be noted that if $\Psi_{\mathbf{r}}$ in (4.7) represents a linear transform, it can be represented by a matrix. In particular, let

M represents the number of functions (columns) in $\Psi_{\mathbf{r}}$ (this implies that the cardinality of the set of $c(\mathbf{r})$ is also M). The set of coefficients in (4.7) can be considered to belong to a Hilbert space $\mathcal{U}$ with its associated definition of the inner product. Accordingly, $\Psi_{\mathbf{r}}$ can be considered as a map from $\mathcal{U}$ to $\mathcal{H}(\Omega_2)$. If $\Psi_{\mathbf{r}}$ represents a sparsifying transform, the support of the set of coefficients $c(\mathbf{r})$, i.e. the number of values of $c(\mathbf{r})$ which are non-zero, is much smaller than the total number of functions of the transform.

It should be noted that the sparsifying transform can often be overcomplete and in that case it is not proper to use the term "basis". Instead, one should use the term "frame" in such a case. Nevertheless, in order to maintain uniformity, with a slight abuse of notation, we are going to use the term "basis" with the understanding that the set of functions can be overcomplete [161].

The application of the theory of compressed sensing requires an *acquisition basis* in the domain of which the signal is acquired. In particular, instead of acquiring a subsampled version of $s(\mathbf{r})$ directly, it is often required to acquire the data in the domain of a certain linear transform $\Phi_{\mathbf{r}}$. Formally, the signal is acquired in the form of $S(\mathbf{r})$, according to

$$\Phi_{\mathbf{r}}\{s(\mathbf{r})\} = S(\mathbf{r}), \tag{4.8}$$

where one can again consider $S(\mathbf{r})$ to be discrete and belong to a Hilbert space $\mathcal{V}$. In the settings of compressed sensing, the dimension (size) of $S(\mathbf{r})$ is much less than that of $s(\mathbf{r})$. It should also be noted that if Φ_r is a linear transform, it can be represented by a matrix. The theory of CS also requires the acquisition domain $\Phi_{\mathbf{r}}$ and the representation domain $\Psi_{\mathbf{r}}$ to be *incoherent*. Intuitively, this implies that $\Psi_{\mathbf{r}}$ should not be a sparsifying transform for the functions (vectors) in $\Phi_{\mathbf{r}}$. In other words, the functions of $\Phi_{\mathbf{r}}$ cannot be expressed as a linear combination of a small number of functions from $\Psi_{\mathbf{r}}$.

In such a setting, the theory of compressed sensing prescribes reconstruction of $c(\mathbf{r})$ for $\mathbf{r} \in \Omega_1$ according to

$$\begin{aligned} c(\mathbf{r}) &= \arg\min_{c} \|c\|_{\mathcal{U},1} \\ \text{s.t. } &\|\mathcal{A}_{\mathbf{r}}\{c\} - S(\mathbf{r})\|_2 \leq \epsilon \end{aligned} \tag{4.9}$$

where $\mathcal{A}_{\mathbf{r}} = \Psi_{\mathbf{r}}\Phi_{\mathbf{r}} : \mathcal{U} \mapsto \mathcal{V}$ and $\epsilon > 0$ depends on the level of contaminating noise. Under certain conditions on $\Phi_{\mathbf{r}}$ and $\Psi_{\mathbf{r}}$ and the type of signal $s(\mathbf{r})$, the theory of compressed sensing guarantees reasonable reconstruction of $c(\mathbf{r})$ from undersampled measurements $S(\mathbf{r})$ (see Chapter 3). Once $c(\mathbf{r})$ is obtained from all $\mathbf{r} \in \Omega_1$, the signal of interest $s(\mathbf{r})$ can be recovered using (4.7).

The approach of performing compressed sensing for functions belonging to $\mathcal{H}(\Omega_2)$ as opposed to performing over the whole data set offers several practical advantages. However,

it ignores the regularity of the data over Ω_1. It should be noted that the real life data is always contaminated with noise. Consequently, the incorporation of the regularity over Ω_1, which provides additional constraint to the reconstruction problem, is expected to provide a more accurate reconstruction of $c(\mathbf{r})$ with fewer samples. A specific and widely used regularity over Ω_1 follows from the fact that the signal defined on Ω_1 is of bounded variation. The latter can be incorporated within the reconstruction framework by minimizing the total variation seminorm (discussed in Chapter 3) defined on the data over Ω_1. Hence, the reconstruction of the data can be obtained

1. by applying the theory of compressed sensing over Ω_2 using some sparsifying transform over $\mathcal{H}(\Omega_2)$, and
2. by regularizing the solution using the additional information that the signal is of bounded variation over Ω_1.

Let us define $\mathfrak{U}$ and $\mathfrak{V}$ to be the Hilbert spaces corresponding to the coefficients and the data respectively for all $\mathbf{r} \in \Omega_1$. To avoid confusion, we reiterate here that $\mathcal{U}$ and $\mathcal{V}$ are the vector spaces corresponding to the coefficients and the acquired data respectively at *each* spatial location $\mathbf{r} \in \Omega_1$. The notations are summarized in Table 4.1.

Furthermore, let us also assume that one is obtaining K measurements of the function in $\mathcal{H}(\Omega_2)$ corresponding to the points $\mathbf{u}_k \in \Omega_2,\ k = 1, \ldots, K$. For each point $\mathbf{u}_k$, the measured data result in the image $S_k \equiv S(\mathbf{u}_k)$, which can be formally viewed as a mapping from Ω_1 to $\mathbb{R}$. On the other hand, at a given coordinate $\mathbf{r} \in \Omega_1$, one can combine the values $S_1(\mathbf{r}), S_2(\mathbf{r}), \ldots, S_K(\mathbf{r})$ into a column vector

$$S(\mathbf{r}) := [S_1(\mathbf{r}), S_2(\mathbf{r}), \ldots, S_K(\mathbf{r})]^T \in \mathbb{R}^K,$$

which defines a discrete vector space on Ω_1. This vector can then be regarded as a vector of discrete measurements of the associated data $S(\mathbf{u} \,|\, \mathbf{r})$ measured at $\{\mathbf{u}_k\}_{k=1}^K \subset \Omega_2$. It is worth noting that, according to the above notations, the value $S_k(\mathbf{r})$ admits a twofold interpretation, *viz.* either as the k^{th} coordinate of vector $S(\mathbf{r})$ or as the value of image S_k at spatial position $\mathbf{r}$.

Next, we need to define the appropriate norms relevant to the present case. First, we define the ℓ_2-norm of $S \in \mathfrak{V}$ as given by

$$\|S\|_{\mathfrak{V},2} = \Big[\sum_{\mathbf{r} \in \Omega_1} \|S(\mathbf{r})\|_2^2 \Big]^{1/2} = \Big[\sum_{k=1}^{K} \|S_k\|_{\mathbb{F}}^2 \Big]^{1/2}, \tag{4.10}$$

Table 4.1: List of notations

Notations	Explanation	Mapping/formula (if applicable)
$\mathcal{U}$	Vector space corresponding to the representation coefficients at *each* $\mathbf{r} \in \Omega_1$	
$\mathcal{V}$	Vector space corresponding to the acquired data at *each* $\mathbf{r} \in \Omega_1$	
$\mathfrak{U}$	Vector space corresponding to the representation coefficients for *all* $\mathbf{r} \in \Omega_1$	
$\mathfrak{V}$	Vector space corresponding to the acquired data for *all* $\mathbf{r} \in \Omega_1$	
$\Psi_{\mathbf{r}}$	Map the representation coefficients to the signal defined over $\mathcal{H}(\Omega_2)$ at each $\mathbf{r} \in \Omega_1$	$\Psi_{\mathbf{r}} : \mathcal{U} \mapsto \mathcal{H}(\Omega_2)$
$\mathcal{A}_{\mathbf{r}}$	Map the representation coefficients to the acquired data at each $\mathbf{r} \in \Omega_1$	$\mathcal{A}_{\mathbf{r}} : \mathcal{U} \mapsto \mathcal{V}$
Ψ	Map the representation coefficients to the signal defined over $\mathcal{H}(\Omega_2)$ for all $\mathbf{r} \in \Omega_1$	$\Psi : \mathfrak{U} \mapsto \mathfrak{V}$
$\mathcal{A}$	Map the representation coefficients to the acquired data for all $\mathbf{r} \in \Omega_1$	$\mathcal{A} : \mathfrak{U} \mapsto \mathfrak{V}$

where $\|\cdot\|_2$ and $\|\cdot\|_{\mathbb{F}}$ denote the Euclidean vector and Frobenius matrix norms, respectively. Another norm on $\mathfrak{V}$ that is relevant is the total variation (TV) semi-norm which is defined as follows. First, let us define the total variation of the k^{th} component S_k of the field S in a standard manner as

$$\|S_k\|_{\text{TV}} = \sum_{\mathbf{r}\in\Omega_1} \Big[\sum_{\mathbf{p}\in\mathcal{C}(\mathbf{r})} |S_k(\mathbf{r}) - S_k(\mathbf{p})|^2 \Big]^{1/2}, \tag{4.11}$$

where $\mathcal{C}\left(\mathbf{r} = (x_{i_1}, y_{i_2}, z_{i_3})\right) = \{(x_{i_1-1}, y_{i_2}, z_{i_3}), (x_{i_1}, y_{i_2-1}, z_{i_3}), (x_{i_1}, y_{i_2}, z_{i_3-1})\}$ is a 3-neighbourhood (causal) clique of voxel $\mathbf{r}$. Consequently, the TV norm of the discrete vector field S can be defined in terms of the TV-norms of its K components as

$$\|S\|_{\mathfrak{V},\text{TV}} = \Big[\sum_{k=1}^{K} \|S_k\|_{\text{TV}}^{\alpha} \Big]^{1/\alpha}. \tag{4.12}$$

Thus, for example, $\alpha = 2$ was used in the TV-denoising method reported in [162]. In this thesis, we use $\alpha = 1$ because it facilitates obtaining a separable structure in the optimization problem as will be discussed later. This definition of TV has been referred as "channel-by-channel" in [163]. A number of additional definitions of total variation for vector valued functions can be found in the same reference.

For the vector space $\mathfrak{U}$, one needs to define the ℓ_1 norm in order to apply the CS framework. This can be defined as

$$\|c\|_{\mathfrak{U},1} = \sum_{\mathbf{r}\in\Omega_1} \|c(\mathbf{r})\|_1 = \sum_{\mathbf{r}\in\Omega_1} \sum_{k=1}^{M} |c_k(\mathbf{r})|. \tag{4.13}$$

Finally, in order to state the reconstruction problem formally, we define the operator $\mathcal{A} : \mathfrak{U} \mapsto \mathfrak{V}$ which maps the coefficients to the acquired data for all $\mathbf{r} \in \Omega_1$. Subsequently, the reconstruction problem can be stated as

$$\begin{aligned} c = \arg\min_{c} \{ &\|c\|_{\mathfrak{U},1} + \gamma \|\mathcal{A}\{c\}\|_{\mathfrak{V},\text{TV}} \} \\ \text{s.t. } &\|\mathcal{A}\{c\} - S\|_{\mathfrak{V},2} \leq \epsilon \end{aligned} \tag{4.14}$$

where the role of $\gamma > 0$ is to balance the relative influence of the sparse and TV terms in the above cost function, and $\epsilon > 0$, as before, depends on the level of contaminating noise.

The incorporation of the additional information over Ω_1 is expected to produce more accurate reconstruction as compared to the CS framework without the regularization over

Ω_1. However, the coupling of the regularizations defined over Ω_1 and Ω_2 respectively increases the computational complexity of the resulting optimization problem. This defeats the principal purpose of performing the CS reconstruction only on $\mathcal{H}(\Omega_2)$ as opposed to the function defined on $\Omega_1 \times \Omega_2$.

In the next section, we will propose a solution to the optimization problem using BTS which will lead to significant simplifications of the solution. In particular, the advantage of BTS is that it helps in splitting the complex optimization problem so that the problem can be solved separately over Ω_1 and Ω_2. The proposed solution is characterized by

1. accuracy of the solution because of combination of the regularity in both Ω_1 and Ω_2, and
2. simplification of solving the problem separately over Ω_1 and Ω_2 which enables one to make the implementation parallelizable.

4.3 Proposed Solution

The optimization problem in (4.14) can be rewritten in its equivalent Lagrangian form

$$\min_{c} \left\{ \frac{1}{2} \|\mathcal{A}\{c\} - S\|^2_{\mathfrak{V},2} + \lambda \|c\|_{\mathfrak{U},1} + \mu \|\mathcal{A}\{c\}\|_{\mathfrak{V},\mathrm{TV}} \right\} \tag{4.15}$$

for some optimal choice of $\lambda > 0$ and $\mu > 0$.

Before specifying the solution using BTS, let us consider the following two cases.

4.3.1 Sparse-only Reconstruction

When $\mu = 0$, the functional in (4.15) becomes *separable* in the domain of Ω_1 in the sense that, in such a case, an optimal c can be recovered by solving

$$\min_{c(\mathbf{r})} \left\{ \frac{1}{2} \|\mathcal{A}_{\mathbf{r}}\, c(\mathbf{r}) - S(\mathbf{r})\|^2_{\mathcal{V},2} + \lambda \, \|c(\mathbf{r})\|_{\mathcal{U},1} \right\} \tag{4.16}$$

at each $\mathbf{r} \in \Omega_1$ independently. Note that (4.16) can be considered to be a Lagrangian form of the optimization problem (4.9). The solution of (4.16) can be obtained efficiently by a number of fast solvers [164, 135]. In this thesis, we use the method of FISTA [135] discussed in Chapter 3.

It should be emphasized that, while being suboptimal from the viewpoint of spatial-domain regularity, the solution of (4.16) through iterative shrinkage is advantageous in two important practical ways. First, it suggests considerable storage reduction, since the solution of (4.16) involves a thresholding operator which sets to zero the representation coefficients with amplitudes less or equal to λ/ν in absolute value. It makes it possible to use sparse data formats to store and manipulate the representation coefficients. Second, the fact that the estimation of $c(\mathbf{r})$ is performed at each voxel independently suggests a natural way to speed up the overall estimation process though parallel computing on a multicore system.

4.3.2 TV-only Reconstruction

When $\lambda = 0$, the problem in (4.15) is equivalent to the problem of minimizing the TV norm with quadratic data fidelity term. Defining the TV norm over $\mathfrak{V}$ in the separable way as mentioned before enables simultaneously solving K optimization problems of the form

$$\min_{c} \left\{ \frac{1}{2} \| [\mathcal{A}\{c\}]_k - S_k \|_{\mathbb{F}}^2 + \mu \, \| [\mathcal{A}\{c\}]_k \|_{\mathrm{TV}} \right\}, \tag{4.17}$$

where $k = 1, \ldots, K$ and $[\mathcal{A}\{c\}]_k$ denotes the k-th component of the vector field $\mathcal{A}\{c\} \in \mathfrak{V}$. Let $[\mathcal{A}\{c\}]_k$ be denoted by u_k, i.e. $u_k := [\mathcal{A}\{c\}]_k$. Then, reformulated with respect to u_k, the problem (4.17) can be rewritten as

$$\min_{u_k} \left\{ \frac{1}{2} \| u_k - S_k \|_{\mathbb{F}}^2 + \mu \, \| u_k \|_{\mathrm{TV}} \right\}, \tag{4.18}$$

in which case it can be recognized as the problem of TV-denoising of the image S_k [5]. A number of efficient solutions of the optimization problem exists (e.g. [143, 144, 33]) which can be used to solve the problem

Next we are going to use the technique of BTS to solve the problem in (4.15).

4.3.3 Solution using BTS

As mentioned before, directly solving the original problem (4.15) is difficult because of the compound nature of the regularization it involves. The BTS technique allows reducing (4.15) to a simpler form through introduction of an auxiliary variable $u \in \mathfrak{V}$, which can

be viewed as a noise-free version of S. Particularly, using u, (4.15) can be redefined as

$$\min_{c,u} \left\{ \frac{1}{2} \|\mathcal{A}\{c\} - S\|_{\mathfrak{V},2}^2 + \lambda \|c\|_{\mathfrak{U},1} + \mu \|u\|_{\mathfrak{V},\mathrm{TV}} \right\} \quad \text{s.t. } \|\mathcal{A}\{c\} - u\|_{\mathfrak{V},2}^2 = 0. \tag{4.19}$$

Then, starting from an arbitrary $b^0 \in \mathfrak{V}$, the BTS algorithm finds optimal c and u through the following iterations

$$\begin{aligned} \left(c^{t+1}, u^{t+1}\right) = \arg\min_{c,u} \Big\{ & \frac{1}{2} \|\mathcal{A}\{c\} - S\|_{\mathfrak{V},2}^2 + \lambda \|c\|_{\mathfrak{U},1} \\ & + \mu \|u\|_{\mathfrak{V},\mathrm{TV}} + \frac{\gamma}{2} \|u - \mathcal{A}\{c\} - b^t\|_{\mathfrak{V},2}^2 \Big\} \\ b^{t+1} = & b^t + \left(\mathcal{A}\{c^{t+1}\} - u^{t+1}\right), \end{aligned} \tag{4.20}$$

for some $\gamma > 0$. The functional in (4.20) is supposed to be minimized over two variables, i.e. u and c. However, due to the way the ℓ_1 and TV components of this functional have been split, the minimization can now be performed efficiently by iteratively minimizing with respect to u and c separately. The resulting iteration steps are

$$\begin{aligned} &\text{Step 1: } c^{t+1} = \arg\min_{c} \left\{ \frac{1}{2} \|\mathcal{A}\{c\} - S\|_{\mathfrak{V},2}^2 + \frac{\gamma}{2} \|\mathcal{A}\{c\} - (u^t - b^t)\|_{\mathfrak{V},2}^2 + \lambda \|c\|_{\mathfrak{U},1} \right\} \\ &\text{Step 2: } u^{t+1} = \arg\min_{u} \left\{ \frac{\gamma}{2} \|u - (\mathcal{A}\{c^{t+1}\} + b^t)\|_{\mathfrak{V},2}^2 + \mu \|u\|_{\mathfrak{V},\mathrm{TV}} \right\}. \end{aligned} \tag{4.21}$$

Note that the functional at Step 1 contains two quadratic terms which can be combined together to result in

$$\text{Step 1: } c^{t+1} = \arg\min_{c} \left\{ \frac{1+\gamma}{2} \left\| \mathcal{A}\{c\} - \frac{S + \gamma\,(u^t - b^t)}{1+\gamma} \right\|_{\mathfrak{V},2}^2 + \lambda \|c\|_{\mathfrak{U},1} \right\}. \tag{4.22}$$

To cause a substantial reduction in the value of the cost functional in (4.20), Step 1 and Step 2 should be applied recursively for a predefined number of times before b^t is updated according to (4.20). It was argued in [33], however, that the extra precision gained through such a repetitive application of Step 1 and Step 2 is likely to be "wasted" when b^t is updated. Consequently, it was suggested in [33] to perform these steps only once per iteration cycle. The convergence of the algorithm is guaranteed by the Eckstein-Bertsekas theorem [37] (see also Theorem 3.1 in [31]).

The final algorithm is summarized below. Lines 3-4 of Algorithm 1 correspond to Step 1 in (4.21) as given by (4.22), while lines 5-6 correspond to Step 2.

Algorithm 3 BTS algorithm for sparse-TV reconstruction of HARDI signals

1: $b \Leftarrow 0$, $u \Leftarrow S$
2: **while** "c keeps changing" **do**
3: $\quad d \Leftarrow (1+\gamma)^{-1}(S + \gamma(u-b))$
4: $\quad c \Leftarrow \arg\min_c \left\{ \frac{1}{2}\|\mathcal{A}\{c\} - d\|^2_{\mathfrak{V},2} + \frac{\lambda}{1+\gamma}\|c\|_{\mathfrak{U},1} \right\}$
5: $\quad d \Leftarrow \mathcal{A}\{c\} + b$
6: $\quad u \Leftarrow \arg\min_u \left\{ \frac{1}{2}\|u - d\|^2_{\mathfrak{V},2} + \frac{\mu}{\gamma}\|u\|_{\mathfrak{V},\mathrm{TV}} \right\}$
7: $\quad b \Leftarrow b + (\mathcal{A}\{c\} - u)$
8: **end while**

4.4 Discussion on Computational Complexity

The proposed solution has a number of computational advantages compared to the direct solution of (4.15).

1. The optimization problem in line 4 of Algorithm 1 is separable in $\mathbf{r} \in \Omega_1$. This optimization, therefore, can be performed at each point $\mathbf{r}$ independently as discussed in subsection 4.3.1. Moreover, the optimization problem in line 6 is separable for different $k = 1, 2, \ldots, K$, and hence it amounts to applying TV-denoising to each of the K components of u independently as discussed in subsection 4.3.2. Since the optimization problems can be solved separately for each point in Ω_1 or Ω_2, the size of the optimization variable as well as the data for each of the problems in line 4 or in line 6 is significantly less than the original problem in (4.15).

2. The steps in line 4 and 6 are parallelizable and hence, one can take advantage of multi-core processing to have efficient implementation of the steps.

3. Each of the problems, namely minimization of quadratic norm with ℓ_1 norm or TV regularizer, is well known and their fast solutions exist in the literature. For example, line 4 for each point in Ω_1 can be solved using the fast techniques of [164, 135] whereas line 6 for each point in Ω_2 can be solved using the fast solvers proposed in [143, 144, 33]. It should be noted that solving the problem in (4.15) directly requires the use gradient based methods which are known to converge slowly.

Next, we demonstrate the applicability of the proposed framework to the problem of reconstruction of HARDI data from its sub-critical measurements. Towards this end, let us first introduce the acquisition principles as well as the data formation model of HARDI.

4.5 HARDI Acquisition and Data Formation Model

Magnetic Resonance Imaging (MRI) is an established gold standard in diagnostic imaging which is capable of producing clear and very detailed contrast of the biological tissues and internal organs. MRI employs neither ionizing radiation, nor excessive levels of electromagnetic energy, and hence it is known to be practically harmless to the human body which makes it applicable to a diverse group of patients, including infants, high-risk pregnant women and elders.

The conventional MRI, however, is incapable of delineating the morphological structure of the white matter in the brain. The white matter within the brain is known to contain the neural fibre tracts. Through reconstructing the pattern of connectivity of the neural tracts in both healthy and diseased subjects, it is therefore possible to obtain an abundance of valuable diagnostic information that could 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 [113]. It is only with the advent of diffusion MRI (dMRI) that scientists have been able to perform quantitative measurements of the diffusivity of the white matter, based on which its structural delineation has become possible [107, 108, 109, 110, 111, 112, 113].

The principle of diffusion MRI is based on the diffusion (random Brownian micromovements) of water molecules. The underlying assumption for the structural delineation is that water molecules tend to diffuse more freely along the main direction of the myelinated axons of nerve cells. The amount of diffusion is estimated by measuring the average movement of the water molecules in a particular direction through the application of a pair of *diffusion-encoding magnetic gradients*. In particular, the amount of diffusion in the particular direction manifests itself as an attenuation of the signal and is assumed to follow a model

$$s = s_0\, e^{-bD} \tag{4.23}$$

where s_0 is the signal intensity without the application of diffusion weighting (conventional MRI), s is the signal with the gradient application, D is the apparent diffusion coefficient in the direction of the applied gradient, and b (commonly known as *b-factor*) is a function of the gradient amplitude, the application time of the gradients and the time between the two gradients [165, Eq. 3.18]. The equation in (4.23) indicates that higher the diffusion coefficient (or the amount of diffusion), the larger the signal loss. Furthermore, the greater the b-factor, the diffusion signal is of higher resolution. However, a greater b-factor is associated with lower signal to noise ratio which results in a trade-off between resolution and SNR.

A very specific and important type of diffusion MRI technique is HARDI. The importance of the technique arise from the fact that it is capable of detecting multiple diffusion flows (or, equivalently, neural fibre tracts) within a given voxel [116, 120, 121, 118, 123, 124, 119, 125]. HARDI acquires diffusion data by sampling a spherical shell at a predefined number of points. Subsequently the acquired HARDI signals can be used to recover their associated *orientation distribution functions* (ODFs), which quantify the multimodal directivity of cerebral diffusion at different spatial locations. In particular, ODFs are functions whose modes are likely to coincide with the direction of local diffusion flows [123]. Yet the above advantage of HARDI over the other dMRI techniques does not come free of charge. Specifically, a standard acquisition of HARDI data requires using about 60-100 diffusion-encoding gradients which imposes severe practical constraints on its use for clinical purposes.

In this thesis, we use a general diffusion model in which each voxel within a region of interest (ROI) is allowed to support more than one fibre tract. Under some reasonable assumptions [121], the diffusion signal $s(\mathbf{u}\,|\,\mathbf{r})$ originating from a voxel with spatial coordinate $\mathbf{r} \in \mathbb{R}^3$ containing $M(\mathbf{r})$ fibres can be modelled as [121, 123, 166]

$$s(\mathbf{u}\,|\,\mathbf{r}) = s_0(\mathbf{r}) \sum_{i=1}^{M(\mathbf{r})} \alpha_i(\mathbf{r}) \exp\left\{-b\,(\mathbf{u}^T D_i(\mathbf{r})\,\mathbf{u})\right\}, \tag{4.24}$$

where $\mathbf{u} \in \mathbb{S}^2$ is the direction of the magnetic gradient (i.e. the direction in which the amount of diffusion is sought), s_0 denotes the diffusion signal obtained in the absence of diffusion encoding gradient and is often called the $b0$ image (i.e. corresponding to $b = 0$), $\alpha_i(\mathbf{r}) > 0$ are positive weights obeying $\sum_{i=1}^{M(\mathbf{r})} \alpha_i(\mathbf{r}) = 1$, and $\{D_i(\mathbf{r})\}_{i=1}^{M(\mathbf{r})}$ are 3×3 diffusion tensors associated with the $M(\mathbf{r})$ neural fibres passing through the coordinate $\mathbf{r}$ and characterizing the direction of diffusion. It should be noted that when normalized by its associated b_0 image $s_0(\mathbf{r})$, the value of $s(\mathbf{u} \mid \mathbf{r})$ quantifies the attenuation of MR readout caused by the diffusion of water molecules in the direction of $\mathbf{u}$ through the spatial coordinate $\mathbf{r}$. The associated ODF corresponding to the HARDI signal can be obtained by taking the Funk Radon transform (FRT) [123] of the signal. A typical HARDI signal with 3 fibres along with its associated ODF is shown in Fig. 4.1. The spherical functions are visualized by means of 3-D surface plots. Such a plot tends to project away from the origin of $\mathbb{R}^3$ in the directions along which a spherical function is maximized, while passing near the origin in the directions where the function approaches zero.

As mentioned before, the HARDI data requires a large number of samples which translates into substantially longer acquisition times. This in turn entails a higher probability for the patient to exercise involuntary motion (typically caused by fatigue and/or stress

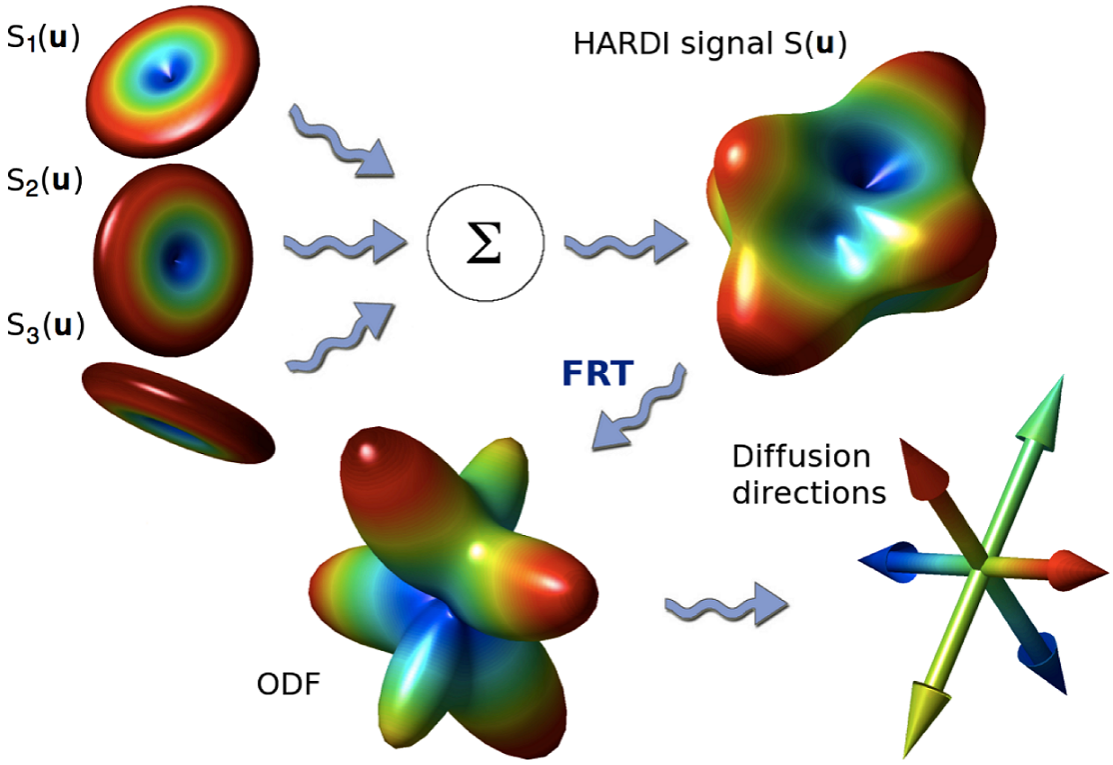


Figure 4.1: HARDI data along with associated ODF

related tremors, swallowing, uncontrollable sighing or coughing), which severely affects the quality of dMRI data. This problem has been addressed in the present chapter using the framework mentioned above. In other words, it has been suggested to acquire fewer samples and subsequently use the theory of compressed sensing to reconstruct the signal while the computational complexity has been significantly reduced by using BTS.

4.5.1 CS Problem Formulation for HARDI

In order to discretize $\Omega_1 \subset \mathbb{R}^3$, it is reasonable to restrict $\mathbf{r} \in \Omega_1$ to belong to a finite rectangular lattice $L_x \times L_y \times L_z$, i.e. $\Omega_1 \equiv \mathbb{R}^{L_x \times L_y \times L_z}$. The application of the theory of compressed sensing requires the HARDI signals to be sparsely represented by a linear transform (representation basis $\Psi_{\mathbf{r}}$). Once available, the signals can be linearly expanded in terms of a relatively small number of functions $\{\psi_m(\mathbf{u})\}_{m=1}^{M}$ of $\Psi_{\mathbf{r}}$. Formally, the signal

$s(\mathbf{u} \mid \mathbf{r})$ can be expanded as

$$s(\mathbf{u} \mid \mathbf{r}) = s_0(\mathbf{r}) \sum_{m=1}^{M} c_m(\mathbf{r}) \psi_m(\mathbf{u}) \tag{4.25}$$

with $c(\mathbf{r}) = [c_1(\mathbf{r}), c_2(\mathbf{r}), \ldots, c_M(\mathbf{r})]^T \in \mathcal{U} \equiv \mathbb{R}^M$ being a vector of representation coefficients which, in general, is a function of the spatial coordinate $\mathbf{r}$.

Next, we need to specify the acquisition basis $\Phi_{\mathbf{r}}$. In the case of HARDI acquisition, the acquisition functions (vectors) of $\Phi_{\mathbf{r}}$ in (4.8) are Dirac deltas as given by

$$\phi_k(\mathbf{u}) = \delta(1 - \mathbf{u} \cdot \mathbf{u}_k), \quad k = 1, 2, \ldots, K \tag{4.26}$$

where it is assumed that we are sampling $\Omega_2 = \mathbb{S}^2$ at K predefined directions $\{\mathbf{u}_k\}_{k=1}^K$. As before, we denote the acquired data after projection on the acquisition basis $\Phi_{\mathbf{r}}$ as $S(\mathbf{r})$ and the corresponding vector space as $\mathcal{V} \equiv \mathbb{R}^K$. Consequently, each sample $S(\mathbf{u}_k|\mathbf{r}_0)$ (for a fixed $\mathbf{r}_0$) of $S(\mathbf{r}_0)$ can be represented as an inner product of $s(\mathbf{u}|\mathbf{r}_0)$ with sampling functions $\{\phi_k\}_{k=1}^K$ as given by

$$S(\mathbf{u}_k|\mathbf{r}_0) = \int_{\mathbb{S}^2} s(\mathbf{u}|\mathbf{r}_0)\, \delta(1 - \mathbf{u} \cdot \mathbf{u}_k)\, d\eta(\mathbf{u}), \quad \text{with } k = 1, 2, \ldots, K \tag{4.27}$$

with $d\eta$ being the standard rotation invariant measure on $\mathbb{S}^2$. In this case, for each $\mathbf{u}_k$, MR measurements are acquired in the form of a diffusion-encoded image $S_k(\mathbf{r}) := S(\mathbf{u}_k \,|\, \mathbf{r})$. As a result, a typical dMRI data set consists of a collection of such diffusion-encoded images $\{S_k(\mathbf{r})\}_{k=1}^K$, whose size K determines the accuracy with which the directions of local diffusion flows can be estimated.

Next, we define the vector spaces $\mathfrak{V}$ and $\mathfrak{U}$ which correspond to the acquired data and the representation coefficients respectively for all $\mathbf{r} \in \Omega_1$. In particular, in order to represent $S \in \mathfrak{V}$ efficiently, it is convenient to look at it as a four dimensional signal whose first three dimensions correspond to the spatial coordinates and the last dimension corresponds to the diffusion encoding direction. Formally,

$$S \in \mathfrak{V} = \mathbb{R}^{L_x \times L_y \times L_z \times K} \tag{4.28}$$

Similarly, the representation coefficients $c \in \mathfrak{U}$, which correspond to the HARDI signals composing s, can be agglomerated into a 4-D array, i.e., $c \in \mathfrak{U} = \mathbb{R}^{L_x \times L_y \times L_z \times M}$, where M denotes the number of functions in $\Psi_{\mathbf{r}}$. Our problem is to recover the representation coefficients $c \in \mathfrak{U}$.

A number of facts render the problem of recovering the representation coefficients non-trivial. Firstly, the HARDI signal(s) S should be expected to be contaminated by considerable levels of measurement noise. In addition, the set of representation coefficients $\{\psi_m\}_{m=1}^M$ can be overcomplete, which suggests that $\dim[\text{Span}\{\psi_m\}_{m=1}^M] < M$. A practical consequence of this fact is that the definition of coefficients $c(\mathbf{r})$ in (4.25) is, in general, not unique. Finally, this non-uniqueness is further aggravated by the fact that $c(\mathbf{r})$ will have to be recovered from an under-sampled set of diffusion measurements, in which case $K \ll M$.

4.5.2 CS based Reconstruction Framework of HARDI

In the case of HARDI CS, as before, $\mathcal{A} : \mathfrak{U} \mapsto \mathfrak{V} : c \mapsto \mathcal{A}\{c\}$ is a linear map whose action is to map the representation coefficients c to their corresponding HARDI signals. Also, $\mathcal{A}_{\mathbf{r}} : \mathbb{R}^M \mapsto \mathbb{R}^K : c(\mathbf{r}) \mapsto \mathcal{A}_{\mathbf{r}}\{c(\mathbf{r})\}$ is a linear operator which maps the representation coefficients at a specific voxel to its corresponding HARDI signal. Using the definitions of the vector spaces and the associated norms given in Section 4.2, we can now formally define the optimization problem which can be used to estimate the representation coefficients. In fact, using the operator $\mathcal{A}$, one can define the HARDI data formation model as

$$S = s_0 \cdot \mathcal{A}\{c\} + e, \tag{4.29}$$

where $e \in \mathfrak{V}$ accounts for both measurement noise and modelling errors. Consequently, congruent to the problem in (4.14), one can define the optimization problem in the case of HARDI as given by

$$\begin{aligned} \min_c \Big\{ \|c\|_{\mathfrak{U},1} + \gamma \, \|\mathcal{A}\{c\}\|_{\mathfrak{V},\text{TV}} \Big\} \\ \text{s.t. } \|s_0 \cdot \mathcal{A}\{c\} - s\|_{\mathfrak{V},2} \leq \epsilon. \end{aligned} \tag{4.30}$$

In the case of HARDI signal, however, one can incorporate additional constraints. In particular, one can consider the fact that the HARDI signal is positive and the incorporation of this constraint can result in more accurate reconstruction. Consequently, the modified optimization problem can be written as

$$\begin{aligned} \min_c \Big\{ \|c\|_{\mathfrak{U},1} + \gamma \, \|\mathcal{A}\{c\}\|_{\mathfrak{V},\text{TV}} + \imath_P(\mathcal{A}\{c\}) \Big\}, \\ \text{s.t. } \|s_0 \cdot \mathcal{A}\{c\} - S\|_{\mathfrak{V},2} \leq \epsilon, \end{aligned} \tag{4.31}$$

where $\imath_P$ denotes the support function of the positive orthant defined as

$$i_P(S) = \begin{cases} 0, & \text{if } S \succeq 0 \\ \infty, & \text{if } S \prec 0 \end{cases} \tag{4.32}$$

where the $\succeq$ and $\prec$ inequalities would be read as "if all" and "if any" respectively. The optimization problem (4.31) can be rewritten in its equivalent Lagrangian form

$$\min_c \left\{ \frac{1}{2} \|s_0 \cdot \mathcal{A}\{c\} - S\|_{\mathfrak{V},2}^2 + \lambda \|c\|_{\mathfrak{U},1} + \mu \|\mathcal{A}\{c\}\|_{\mathfrak{V},\mathrm{TV}} + i_P(\mathcal{A}\{c\}) \right\}, \tag{4.33}$$

for some optimal values of $\lambda > 0$ and $\mu > 0$ [28].

4.5.3 Solution using BTS

Next, we provide the formulation of BTS for the case of (4.33). It should be noted that the solution steps are different from that in section 4.3 because of the incorporation of the positivity constraint. In particular, using the two auxiliary variables $u \in \mathfrak{V}$ and $v \in \mathfrak{V}$, the problem in (4.33) becomes

$$\begin{aligned} \min_{c,u,v} & \left\{ \frac{1}{2} \|s_0 \cdot \mathcal{A}\{c\} - S\|_{\mathfrak{V},2}^2 + \lambda \|c\|_{\mathfrak{U},1} + \mu \|u\|_{\mathfrak{V},\mathrm{TV}} + i_P(v) \right\}, \\ & \text{s.t. } \|u - \mathcal{A}\{c\}\|_{\mathfrak{V},2}^2 = 0, \quad \|v - \mathcal{A}\{c\}\|_{\mathfrak{V},2}^2 = 0 \end{aligned} \tag{4.34}$$

Then, starting from arbitrary $b_u^0 \in \mathfrak{V}$ and $b_v^0 \in \mathfrak{V}$, (5.11) can be solved iteratively according to

$$\begin{aligned} (c^{t+1}, u^{t+1}, v^{t+1}) = \arg\min_{c,u,v} & \left\{ \frac{1}{2} \|s_0 \cdot \mathcal{A}\{c\} - S\|_{\mathfrak{V},2}^2 + \lambda \|c\|_{\mathfrak{U},1} + \mu \|u\|_{\mathfrak{V},\mathrm{TV}} \right. \\ & \left. + i_P(v) + \frac{\delta_u}{2} \|u - \mathcal{A}\{c\} - b_u^t\|_{\mathfrak{V},2}^2 + \frac{\delta_v}{2} \|v - \mathcal{A}\{c\} - b_v^t\|_{\mathfrak{V},2}^2 \right\}, \end{aligned} \tag{4.35}$$

followed by the update

$$b_u^{t+1} = b_u^t + (\mathcal{A}\{c^{t+1}\} - u^{t+1}), \tag{4.36}$$

$$b_v^{t+1} = b_v^t + (\mathcal{A}\{c^{t+1}\} - v^{t+1}) \tag{4.37}$$

where the superscripts indicate the iteration number. Note that the choice of the parameter $\delta_u, \delta_v > 0$ in (4.35) is arbitrary [33], and in the present work, they are set to be equal to 0.5.

The minimization problem in (4.35) can be solved sequentially as a succession of separate optimizations w.r.t. c, u and v, as given by

$$c^{t+1} = \arg\min_{c} \left\{ \frac{1}{2}\|s_0 \cdot \mathcal{A}\{c\} - S\|^2_{\mathfrak{V},2} + \frac{\delta_u}{2}\|u^t - \mathcal{A}\{c\} - b^t_u\|^2_{\mathfrak{V},2} + \frac{\delta_v}{2}\|v^t - \mathcal{A}\{c\} - b^t_v\|^2_{\mathfrak{V},2} + \lambda\|c\|_{\mathfrak{U},1} \right\}, \tag{4.38}$$

$$u^{t+1} = \arg\min_{u} \left\{ \frac{\delta_u}{2}\|u - \mathcal{A}\{c^{t+1}\} - b^t_u\|^2_{\mathfrak{V},2} + \mu\|u\|_{\mathfrak{V},\mathrm{TV}} \right\} \tag{4.39}$$

$$v^{t+1} = \arg\min_{v} \left\{ \frac{\delta_v}{2}\|v - \mathcal{A}\{c^{t+1}\} - b^t_v\|^2_{\mathfrak{V},2} + i_P(v) \right\} \tag{4.40}$$

Specific solution methods for problems (4.38) and (4.39) and (4.40) are detailed below.

As before, the optimization problem in (4.38) is *separable in the spatial coordinate*, and can be solved in a voxel-by-voxel manner. Specifically, the solution of (4.38) amounts to solving a set of optimization problems, which for each $\mathbf{r}$ have the form

$$c^{t+1}(\mathbf{r}) = \arg\min_{c(\mathbf{r})} \left\{ \frac{1}{2}\|\mathcal{A}_{\mathbf{r}}\{c(\mathbf{r})\} - \hat{S}(\cdot|\mathbf{r})\|^2 + \hat{\lambda}(\mathbf{r})\|c(\mathbf{r})\|_1 \right\}, \tag{4.41}$$

where

$$\hat{\lambda}(\mathbf{r}) = \lambda/(s_0(\mathbf{r})^2 + \delta_u + \delta_v), \tag{4.42}$$

and

$$\hat{S}(\cdot|\mathbf{r}) = \frac{s_0(\mathbf{r})S(\cdot|\mathbf{r}) + \delta_u(u^t(\cdot|\mathbf{r}) - b^t_u(\cdot|\mathbf{r})) + \delta_v(v^t(\cdot|\mathbf{r}) - b^t_v(\cdot|\mathbf{r}))}{s_0(\mathbf{r})^2 + \delta_u + \delta_v}. \tag{4.43}$$

The problem in (4.41) can be solved in a computationally efficient manner by means of the FISTA algorithm [135].

At the same time, minimization in (4.39) is *separable in the diffusion coordinate*, which suggests that it can be accomplished through solving a total of K standard TV-denoising problems with a quadratic data-fidelity term. Such a problem can be solved efficiently using the fixed point algorithm of [143].

Finally, the solution of (4.40) is obtained by simply projecting $\mathcal{A}\{c^{t+1}\} + b^t_v$ on the positive orthant, viz.

$$v^{t+1} = \max\{\mathcal{A}\{c^{t+1}\} + b^t_v, 0\}, \tag{4.44}$$

where the maximum is computed in an element-wise way.

4.6 Results

4.6.1 Technical Details of the Experimental Study

To assess the performance of the proposed algorithm under controllable conditions, experiments with simulated data sets have been performed. In this case, the HARDI signals were generated according to model (4.24) with different values of $M(\mathbf{r})$ and $D_i(\mathbf{r})$, and $s_0(\mathbf{r}) = 1\ \forall \mathbf{r}$. The resulting signals were contaminated by variable levels of Rician noise, giving rise to a set of different SNRs. In this work, we adapt the standard definition of the SNR as

$$\mathrm{SNR} = 20 \log_{10} \left(\frac{\|s\|_{\mathfrak{V},2}}{\|s - \tilde{s}\|_{\mathfrak{V},2}} \right), \tag{4.45}$$

where s and $\tilde{s}$ denote an original signal and its noise-contaminated version, respectively, and the norms are computed as defined by (4.10).

4.6.2 Representation Basis

In this work, we use three representation basis, namely the spherical ridgelets [132], spherical harmonics [167, 119, 125] and the spherical Gaussian function [128, 127].

The first one is the basis of spherical ridgelets which are spherical functions having their $\mathbb{L}_2$-energy compactly supported alongside the great circles of $\mathbb{S}^2$. The fact that makes it a useful tool for analysis of HARDI data is that the ridgelet representation is a multiresolution technique, which possesses an intrinsic ability to deal with a continuum of different diffusion scales. Specifically, HARDI signal can be expanded using a relatively small number of spherical ridgelet functions. For the construction of the spherical ridgelets, following [132], the scaling parameter ρ and the resolution parameter J are set to 0.5 and 1 respectively. It should be noted that the value of $J = 1$ corresponds to three resolution levels. The number of spherical ridgelet orientations were predefined with $m_0 = 4$, resulting in $M_{-1} = 16$, $M_0 = 49$ and $M_1 = 169$ ridgelets spanning the resolution levels $j = -1$, $j = 0$ and $j = 1$, respectively. Thus, the total number of spherical ridgelets used in the reconstruction was equal to 234.

The second choice is the basis of spherical harmonics (SH), which although is not optimized for sparsely representing HARDI signal, is a standard representation basis for functions defined on sphere [167, 119, 125]. In particular, in this work, we use spherical harmonics up to the order 8 inclusive. In the case of a real and symmetric analysis, this SH-basis consists of 45 functions.

Additionally, the representation system proposed in [128, 127] has been exploited in this study as well. This system is formed by applying a set of rotations to a Gaussian kernel of the form $d(\mathbf{u}) = \exp\{-b\,(\mathbf{u}^T D_0\,\mathbf{u})\}$, with b defined as in (4.24) and D_0 equal to a (diagonal) diffusion tensor having a mean diffusivity of 766 mm^2/s and a fractional anisotropy of 0.8. Following [128], the number of rotations (and hence the number of Gaussian basis functions) was set to be equal to 253. For the convenience of referencing, the CS-based reconstruction methods using the spherical ridgelets, the 8th-order spherical harmonics, and the rotated Gaussian kernels will be referred below to as the RDG, SH8 and GSS algorithms, respectively.

To assess the significance of the combination of the spatial regularization with the CS framework, all the above algorithms have been applied with two different values of μ in (4.33), *viz.* $\mu = 0$ and $\mu = 0.05$. In the first of these cases, the spatial regularity is ignored, which leads to the sparse-only reconstruction discussed in Section 4.3.1. In the second case, on the other hand, the spatial regularity is taken into account and the reconstruction is performed by means of BTS of Section 4.5.3.

4.6.3 Performance Metrics

To quantitatively compare the reconstruction results produced by the proposed and references methods for different numbers of sampling directions K and various SNRs, three performance measures were used. The first of the three was the normalized mean-square error (NMSE) defined as

$$\mathrm{NMSE} = \frac{1}{L_x L_y L_z} \sum_{\mathbf{r}\in\Omega_1} \frac{\|s(\mathbf{r}) - \hat{s}(\mathbf{r})\|_2^2}{\|s(\mathbf{r})\|_2^2}, \tag{4.46}$$

with $s(\mathbf{r})$ being a reference HARDI signal corresponding to location $\mathbf{r}$ and $\hat{s}(\mathbf{r})$ being its estimate. Depending on the nature of a specific experiment, the reference signal can be either a simulated signal discretized at 642 spherical points obtained by the 3rd order tessellation of the icosahedron *or* a signal reconstructed using a maximum possible number of diffusion-encoding orientations.

One of the most valuable outcomes of HARDI is in providing an access to the computation of ODFs. The latter can in turn be used to recover the directions of local diffusion flows (or, equivalently, the orientations of their related fibre tracts) using, e.g., the steepest ascent procedure detailed in [132]. Suppose $\mathbf{u}_0$ is the true direction of a diffusion flow and $\hat{\mathbf{u}}$ is its estimate. Then, the angular orientation error δ can be defined (in degrees) as

$$\delta = \frac{180}{\pi} \arccos(\mathbf{u}_0 \cdot \hat{\mathbf{u}}). \tag{4.47}$$

In this work, as a performance measure, we use an *average* angular orientation error which is obtained by averaging the values of δ computed for all "fibres" within a specified Ω_1.

The last performance measure used in this work is the probability P_d of false fibre detection. To define P_d, let $M(\mathbf{r})$ be the true number of fibre tracts passing through voxel $\mathbf{r}$ (as defined by model (4.24)). Also, let $\hat{M}(\mathbf{r})$ be an estimated number of fibres, which is equal to the number of modes (maxima) of the ODF recovered at position $\mathbf{r}$. Then, one can define

$$P_d = \left[\frac{1}{L_x L_y L_z} \sum_{\mathbf{r} \in \Omega_1} \frac{|M(\mathbf{r}) - \hat{M}(\mathbf{r})|}{M(\mathbf{r})}. \right] \cdot 100\%. \tag{4.48}$$

In addition to the quantitative comparison, the reconstruction results will be evaluated through visual comparison as well.

4.6.4 In Silico Experiments

To assess the performance of the proposed and reference methods under controllable conditions, two simulated data sets were used. The first set (referred to below as Phantom #1) had a spatial dimension of 12×12 pixels, and consisted of two "fibres" crossing each other at the right angle as it is shown in the upper row of subplots of Fig. 4.2. In addition, each pixel in the set was assigned an extra diffusion flow in the direction perpendicular to the image plane. As a result, the number of diffusion components $M(\mathbf{r})$ in Phantom #1 varied between 1 and 3. Subsequently, model (4.24) was used to generate corresponding diffusion-encode images $\{s_k\}_{k=1}^K$ for a range of different values of K. Two different values of b, namely $b = 1000$ s/mm^2 and $b = 3000$ s/mm^2 were used for data generation. The diffusion tensors $D_i(\mathbf{r})$ in (4.24) were obtained by applying rotations to a tensor of the form $D_0 = \mathrm{diag}\,([\alpha, \beta, \beta])$, where α and β were equal to $1700 \cdot 10^{-6}$ and $300 \cdot 10^{-6}$, respectively. The mean diffusivity and fractional anisotropy of D_0 are equal to 766 mm^2/s and 0.8, respectively. Thus, the same diffusion tensors were used for data synthesis and for the construction of basis functions in the GSS algorithm, thereby allowing the latter to perform under the best possible conditions.

The lower row of subplots in Fig. 4.2 depicts four examples of the diffusion-encoding images obtained for Phantom #1 before their contamination by Rician noise. One can see that the images are piecewise constant functions, which appears to be in a good agreement with the bounded-variation model suggested by (4.33). However, real images may be more complicated than that. Accordingly, to test the robustness of the proposed regularization scheme, a different *in silico* phantom was designed. Phantom #2 had a spatial dimension of

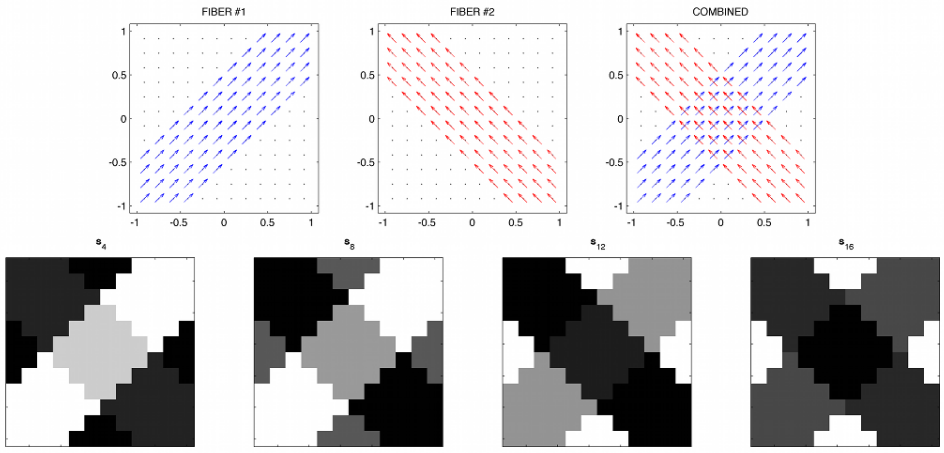


Figure 4.2: Phantom #1: (Upper row of subplots) The orientations of the individual diffusion flows and their combination; (Lower row of subplots) Examples of the resulting (noise-free) diffusion-encoding images corresponding to four different diffusion-encoding directions.

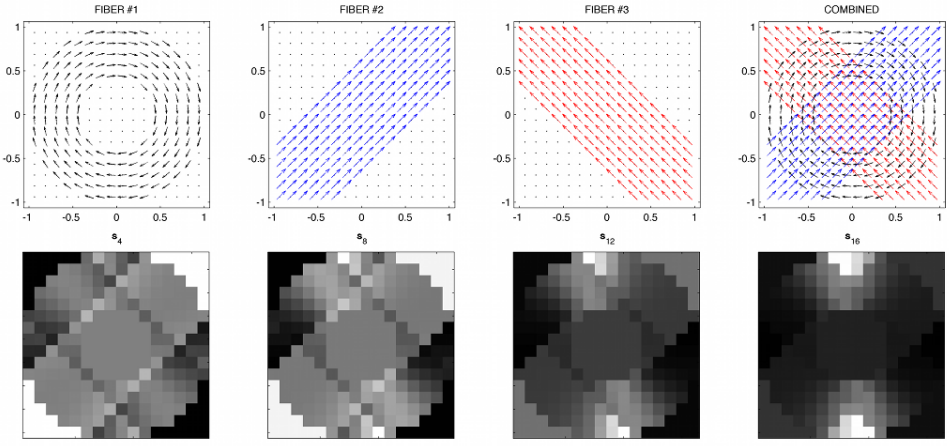


Figure 4.3: Phantom #2: (Upper row of subplots) The orientations of the individual diffusion flows and their combination; (Lower row of subplots) Examples of the resulting (noise-free) diffusion-encoding images corresponding to four different diffusion-encoding directions.

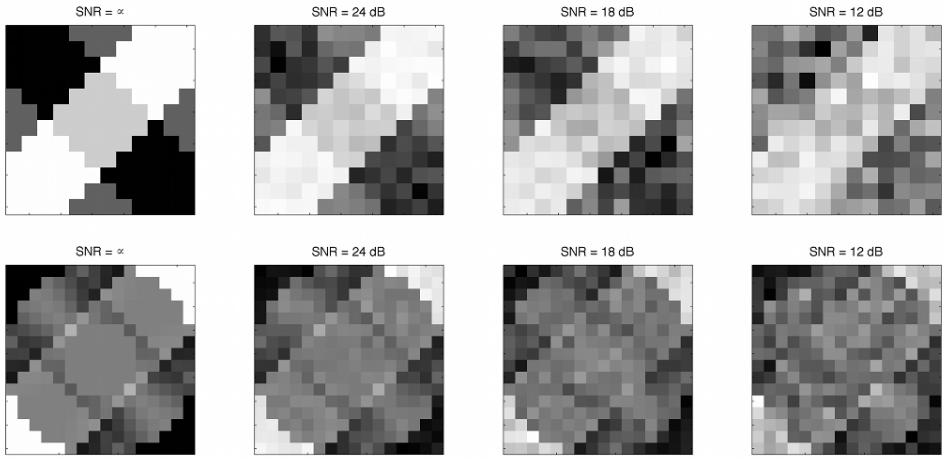


Figure 4.4: (Upper row of subplots) Diffusion-encoding images of Phantom #1 corresponding to $\mathbf{u} = [1, 1, 1]/\sqrt{3}$ and SNR $= \infty$, 24, 18 and 12 dB; (Lower row of subplots) Diffusion-encoding images of Phantom #2 corresponding to the same $\mathbf{u}$ and the same values of SNR.

16×16 pixels and it was obtained through supplementing the configuration of Phantom #1 by an additional circular "fibre" as shown in the upper row of subplots in Fig. 4.3. The lower row of subplots of the figure shows a subset of the resulting diffusion-encoded images, which can be seen to no longer exhibit a piecewise constant behaviour characteristic for Phantom #1.

The simulated diffusion-encoded images were contaminated by three different levels of Rician noise, giving rise to SNR of 24, 18 and 12 dB. Some typical examples of the resulting images are demonstrated in Fig. 4.4, where the upper row of subplots depicts a noise-free version of one of the diffusion-encoded images pertaining to Phantom #1 along with its noise-contaminated counterparts. The lower row of subplots in Fig. 4.4 depicts an analogous set of examples for Phantom #2. Observing the figure, one can see that the SNR values have been chosen so as to cover a range of possible noise scenarios, which could be characterized as moderate-to-severe contamination.

As it was mentioned earlier, in our *in silico* study we compared the performance of three different representation bases, i.e. spherical harmonics (SH8), Gaussian kernels (GSS) and spherical ridgelets (RDG). All the resulting algorithms have been further subdivided into two different types, depending on whether or not the spatial regularization was engaged. Thus, in the absence of the spatial regularization (corresponding to $\mu = 0$), the reconstruction has been performed on a voxel-by-voxel basis, as detailed in Section 4.3.1. For

the convenience of referencing, the corresponding algorithms will be referred to below as SH8-CS, GSS-CS, and RDG-CS. In the case of $\mu > 0$, the estimation has been carried out using the method of BTS as provided in Section 4.5.3. The corresponding algorithms will be referred below as SH8-TV, GSS-TV, and RDG-TV.

The upper subplot of Fig. 4.5 shows the original field of ODFs of Phantom #1 (corresponding to $b = 3000$ s/mm^2), which has been computed based on Tuch's approximation [123] (i.e. by applying the Funk-Radon transform to the diffusion signals). At the same time, the middle row of subplots of Fig. 4.5 shows the ODFs recovered by (from left to right) SH8-CS, GSS-CS and RDG-CS with $K = 16$ and SNR $= 18$ dB. One can see that the inability of the SH basis to sparsely represent HARDI signals results in a poor performance of SH8-CS. A better result is obtained with GSS-CS, which uses a basis of rotated Gaussian kernels, and therefore has a potential to represent the HARDI signals in a sparse manner. Unfortunately, the excessive correlation between the Gaussian basis functions adversely affects the ability of this method to withstand the effect of noise. Consequently, the reconstruction obtained using GSS-CS suffers from sizeable errors. The RDG-CS method, on the other hand, provides an estimation result of a much higher quality, albeit some inaccuracies are still noticeable in the central part of the phantom. The reconstruction accuracy improves dramatically when the spatial regularization is "switched on", as it is demonstrated by the bottom row of subplots in Fig. 4.5. Specifically, while SH8-TV is still unable to provide a valuable reconstruction, the estimates obtained using GSS-TV and RDG-TV represent correctly the "flow structure" of Phantom #1. Moreover, among the latter two methods, RDG-TV is clearly the best performer, resulting in a close-to-ideal recovery of the original ODFs. The superiority of RDG-TV over the alternative methods is further evident in the results presented by Fig. 4.6, which depict the reconstructions obtained for Phantom #2 (with the same values of b, K and SNR as above).

In general, the reconstruction results obtained using SH8-CS and SH8-TV have been observed to be of a lower quality in comparison to the other methods under consideration. For this reason, in what follows, only the GSS and RDG methods are compared. Thus, Fig. 4.7 contrasts the performances of GSS-CS, GSS-TV, RDG-CS and RDG-TV in terms of the NMSE criterion. One can see that the best performance here is attained by the RDG-TV algorithm, which results in the smallest values of NMSE for both phantoms and for all the tested values of b, SNR and K. It is also interesting to note that the incorporation of spatial regularization allows GSS-TV to outperform RDG-CS, with the effect of the regularization becoming more pronounced at lower SNRs. On the whole, all the NMSE curves demonstrate an expected behaviour, with the error values increasing proportionally with a decrease in SNR, while going down with an increase in the number of diffusion-encoding gradients K. However, as opposed to the others, the NMSE curves

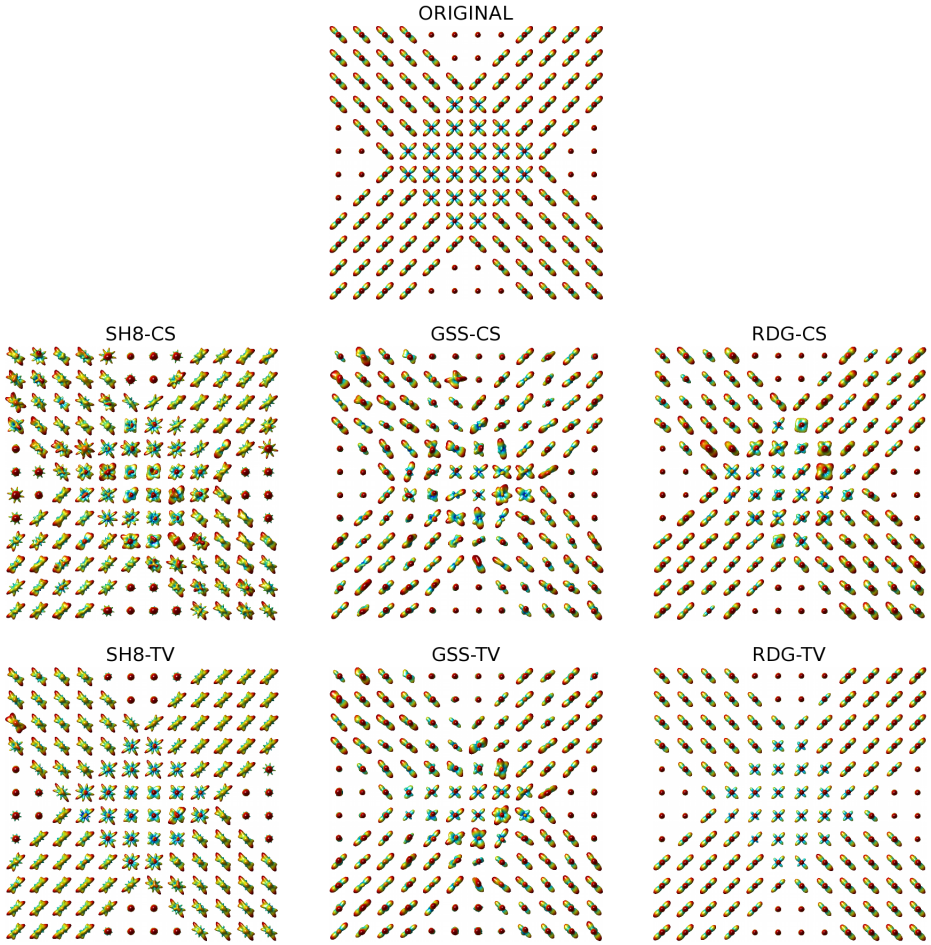


Figure 4.5: (Upper subplot) Original ODFs of Phantom #1; (Middle row of subplots) The ODFs recovered by the SH8-CS, GSS-CS, and RDG-CS algorithms, respectively; (Bottom row of subplots) The ODFs recovered by the SH8-TV, GSS-TV, and RDG-TV algorithms, respectively.

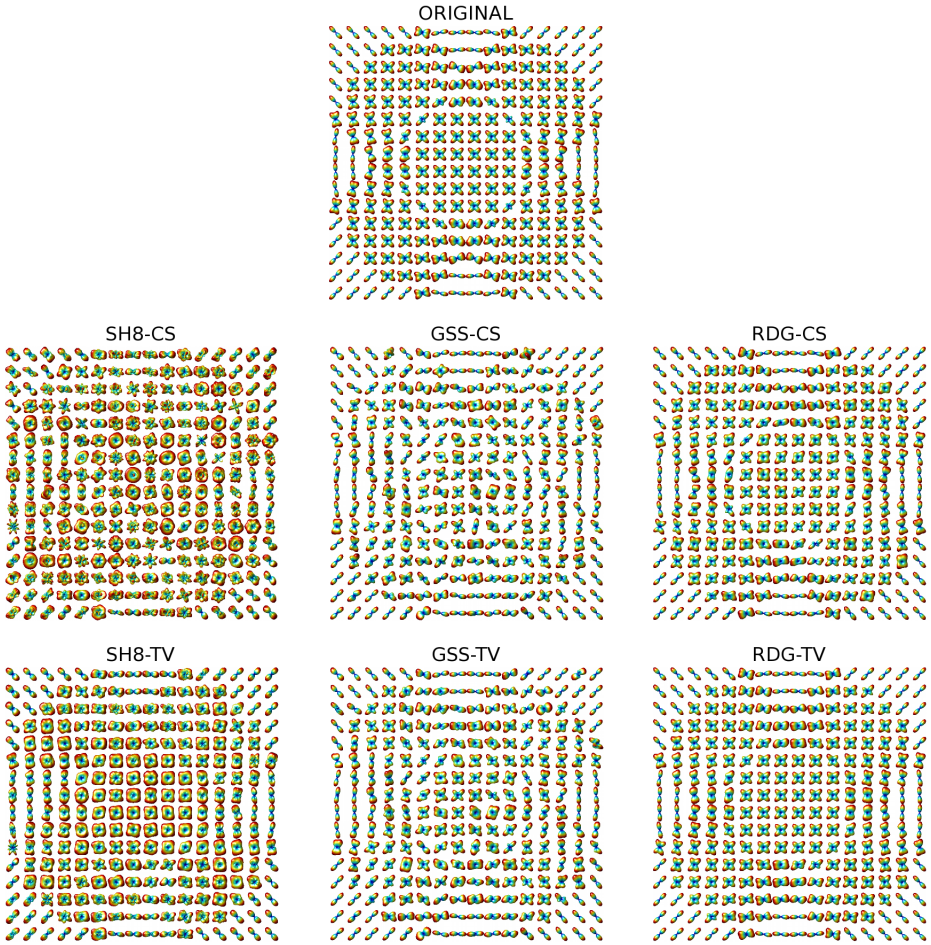


Figure 4.6: (Upper subplot) Original ODFs of Phantom #2; (Middle row of subplots) The ODFs recovered by the SH8-CS, GSS-CS, and RDG-CS algorithms, respectively; (Bottom row of subplots) The ODFs recovered by the SH8-TV, GSS-TV, and RDG-TV algorithms, respectively.

obtained with RDG-TV are characterized by a relatively low rate of convergence, which indicates a reduced sensitivity of RDG-TV to the value of K.

The above algorithms have also been compared in terms of the angular error (4.47). The results of this comparison are summarized in Fig. 4.8, which again indicates that the most accurate reconstruction is obtained using the RDG-TV method. In general, the angular error tends to grow as SNR decreases and to converge to a minimum as K increases. As opposed to the case of NMSE, however, there is an additional dependency of the angular error on the type of a phantom in use as well as on the b-value. In particular, the errors obtained for Phantom #2 are (on average) greater than those obtained for Phantom #1. This discrepancy is rooted in the fact that Phantom #2 has a more complex "fibre structure" as compared to Phantom #1. Specifically, while the "fibers" of Phantom #1 are designed to cross each other at the right angle, the "fibres" of Phantom #2 are allowed to decussate at much smaller angles, which makes them much harder to resolve. Moreover, this effect becomes more noticeable with a decrease in the b-value, which reduces the resolution of q-ball imaging. Finally, we notice that, on average, GSS-TV performs better than RDG-CS (though still worse than RDG-TV), which justifies the value of spatial regularization.

The comparison in terms of the rate of false fibre detection P_d (4.48) was last in the line of our *in silico* performance tests; its results are shown in Fig. 4.9. One can see that, in the case of Phantom #1, RDG-TV yields a virtually zero false detection rate for both values of b, whereas the other methods result in considerably higher values of P_d (mainly due to the detection of spurious local maxima in the estimated ODFs). The situation is different for Phantom #2, where all the compared methods yield sizeable errors (especially for $b = 1000$ s/mm^2). However, in comparative terms, the most accurate reconstruction is still obtained by means of the proposed RDG-TV algorithm.

4.6.5 In Vivo Results

As the next validation step, experiments with real HARDI data were carried out. The proposed algorithm was tested on human brain scans acquired on a 3-Tesla GE system using an echo planar imaging (EPI) diffusion-weighted image sequence. A double echo option was used to suppress eddy-current related distortions. To improve the spatial resolution of EPI, an eight channel coil was used to perform parallel imaging by means of the ASSET technique with a speed-up factor of 2. The data were acquired using 51 gradient directions (quasi-uniformly distributed over the northern hemisphere) with $b = 1000$ s/mm^2. In addition, eight baseline ($b0$) scans were acquired, averaged and used for normalization. The following scanning parameters were used: repetition time (TR) = 17000 ms, echo

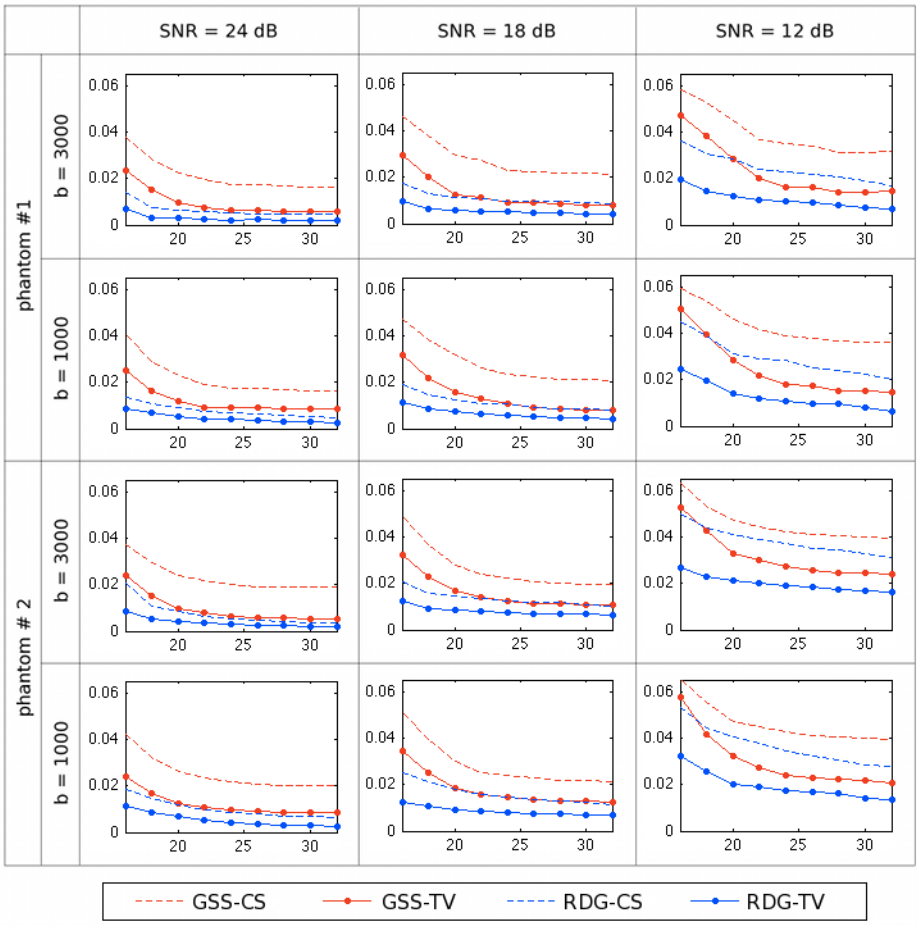


Figure 4.7: NMSE obtained using the compared methods for different phantoms, SNRs and b-values.

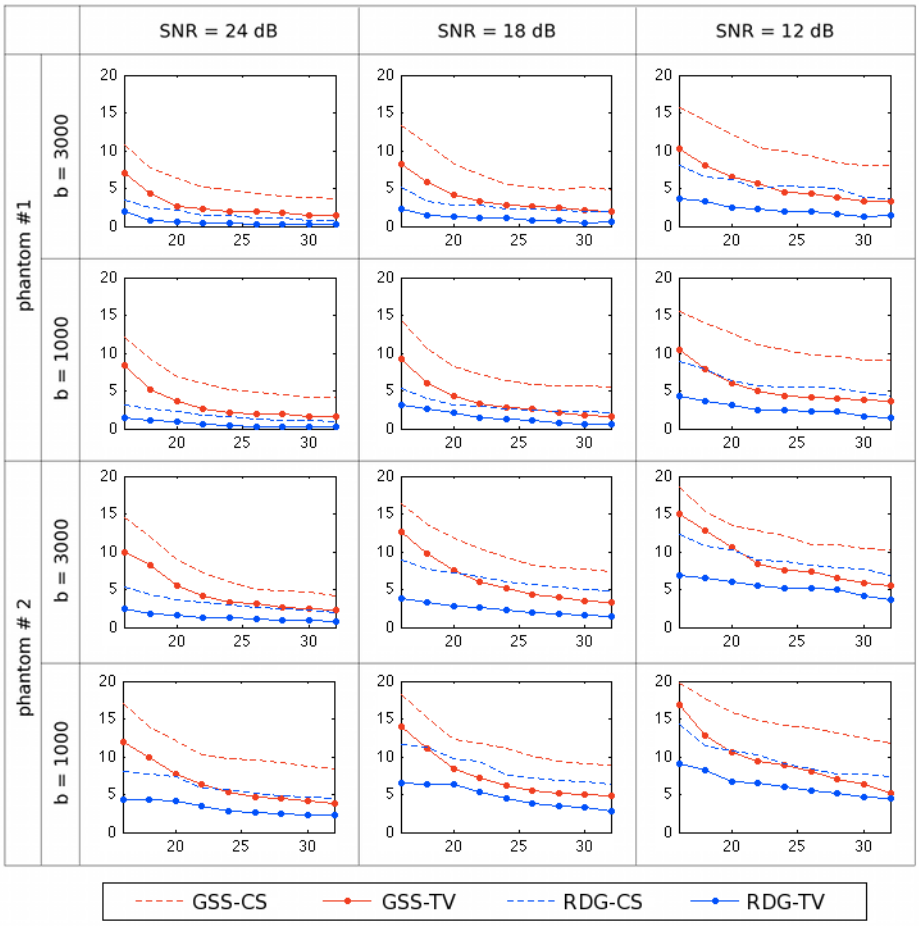


Figure 4.8: Average angular error δ obtained using the compared methods for different phantoms, SNRs and b-values.

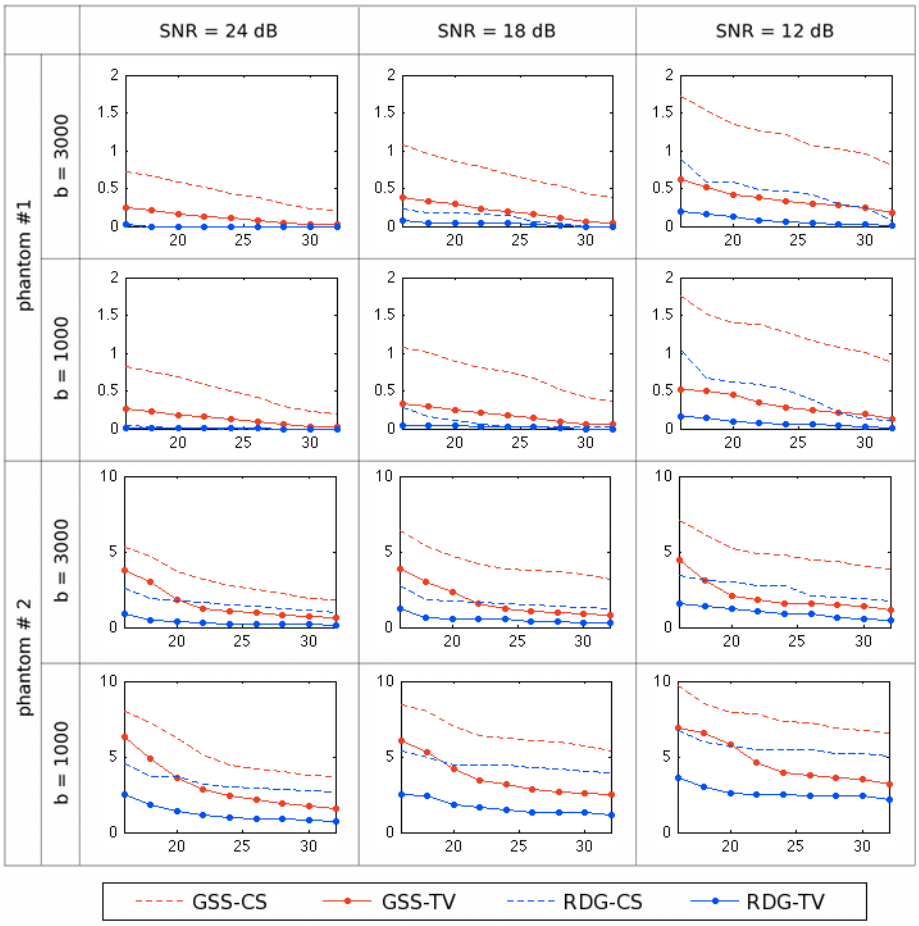


Figure 4.9: The rate of false fibre detection P_d obtained using the compared methods for different phantoms, SNRs and b-values.

time (TE) = 78 ms, field of view (FOV) = 24 cm, 144 × 144 encoding steps, and 1.7 mm slice thickness. All scans had 85 axial slices parallel to the AC-PC line covering the whole brain.

The main question addressed through the *in vivo* experiments has been whether or not it is possible to supersede the spatial regularization by pre-filtering of HARDI signals. To this end, the RDG-CS algorithm was applied first to the HARDI data containing the full set of $K = 51$ diffusion gradients. (Note that such *dense* reconstruction is analogous to the one reported in [132], where the latter is shown to outperform the SH-based estimation [125].) The resulting ODFs have been used as a fiducial against which different reconstruction results were compared.

As the next step, three different subsets of 16, 24 and 32 spherical points were composed out of the original set of 51 diffusion gradients. Within each of these subsets, their corresponding points were chosen so as to result in a quasi-uniform coverage of the northern hemisphere. Accordingly, the HARDI data were rearranged into three data sets of size $144 \times 144 \times 85 \times 16$, $144 \times 144 \times 85 \times 24$ and $144 \times 144 \times 85 \times 32$ to emulate compressed sensing data acquisition. The above sets were used to assess the performance of different reconstruction methods. Unfortunately, we have not succeeded to find conditions under which the SH8 and GSS algorithms would provide stable reconstruction results (either with or without pre-filtering). For this reason, only the RDG-CS and RDG-TV algorithms are compared below.

Table 4.2: NMSE computed between the dense and CS-based reconstructions obtained with RDG-CS and RDG-TV

	Pertaining to Fig. 4.10			Pertaining to Fig. 4.11		
	$K = 16$	$K = 24$	$K = 32$	$K = 16$	$K = 24$	$K = 32$
RDG-CS	0.097	0.064	0.043	0.091	0.053	0.037
RDG-TV	0.022	0.011	0.003	0.018	0.009	0.002

The upper row of subplots in Fig. 4.10 show the generalized anisotropy (GA) [123] image of a coronal cross-section of the brain along with the reference field of ODFs corresponding to the region indicated by the yellow rectangular. Anatomically, this region is expected to contain the fibre bundles of corona radiata as well as those of superior longitudinal and arcuate fasciculi. The middle row of subplots in the same figure depict the ODFs reconstructed by RDG-CS using $K = 16, 24$ and 32 diffusion gradients. One can see that the quality of reconstruction progressively improves as K increases. It is important to note that, before applying the RDG-CS algorithm, the diffusion-encoded images had been pre-processed by a TV filter to reduce the effect of measurement noises on the estimation

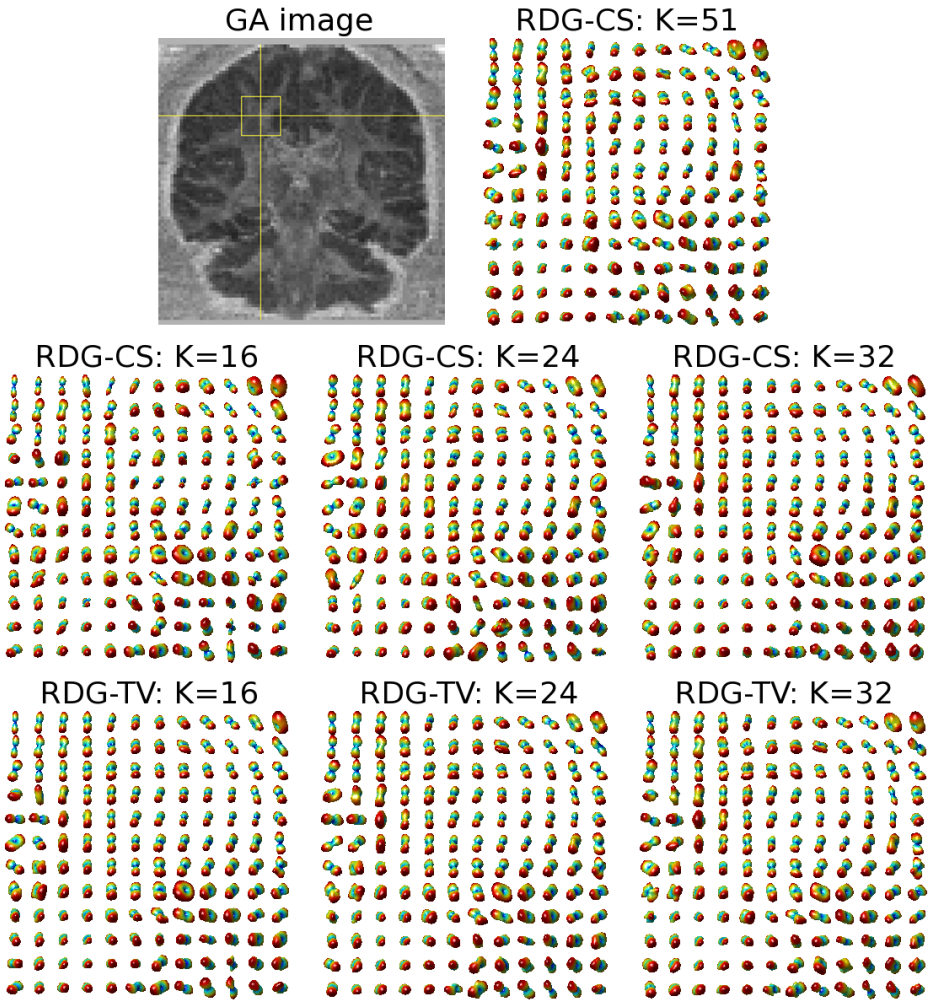


Figure 4.10: (*Upper row of subplots*) A coronal GA image and the ODF field of the indicated region recovered by RDG-CS with $K = 51$; (*Middle row of subplots*) Estimated ODF fields obtained using RDG-CS with $K = 16$, $K = 24$ and $K = 32$; (*Bottom row of subplots*) Estimated ODF fields obtained using RDG-TV with $K = 16$, $K = 24$ and $K = 32$.

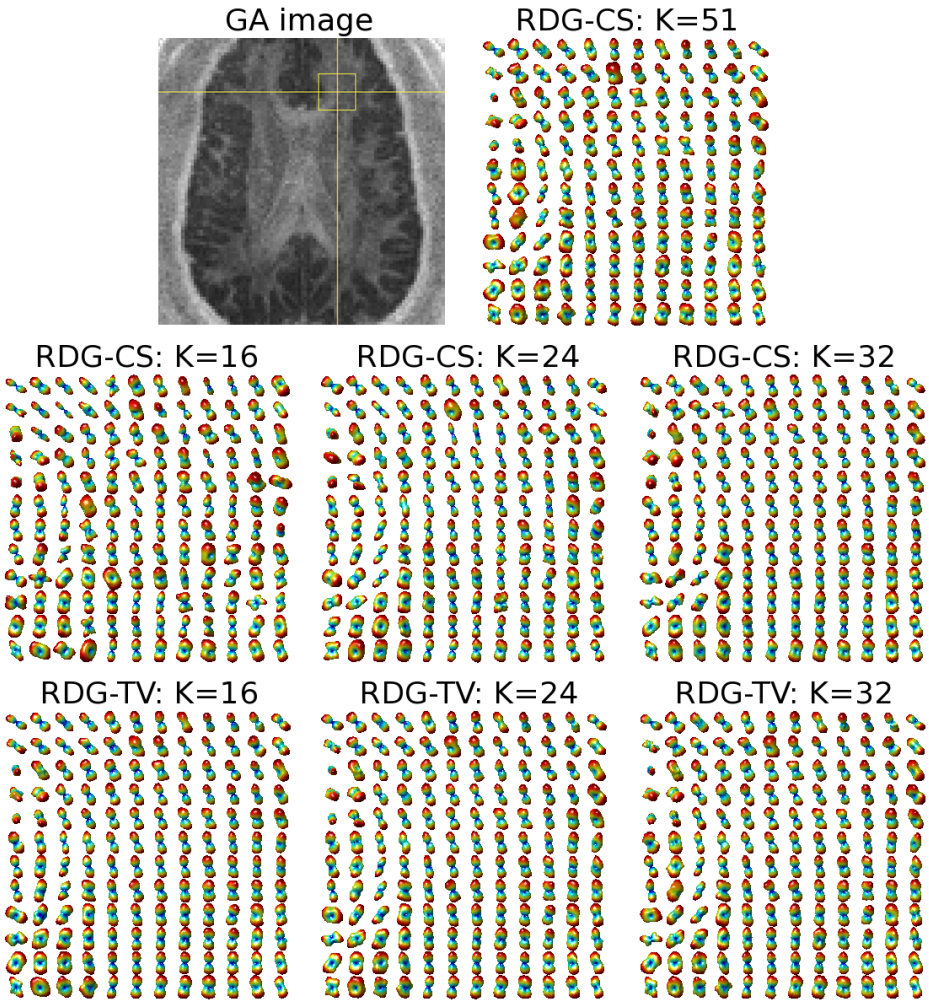


Figure 4.11: (*Upper row of subplots*) An axial GA image and the ODF field of the indicated region recovered by RDG-CS with $K = 51$; (*Middle row of subplots*) Estimated ODF fields obtained using RDG-CS with $K = 16$, $K = 24$ and $K = 32$; (*Bottom row of subplots*) Estimated ODF fields obtained using RDG-TV with $K = 16$, $K = 24$ and $K = 32$.

result. However, this pre-processing appears to be not nearly as effective as the spatial regularization of the RDG-TV algorithm, whose reconstruction results are shown in the bottom row of subplots in Fig. 4.10. The above conclusion is further supported by an additional example of Fig. 4.11, which shows the reconstructions pertaining to the indicated area within an axial cross-section of the brain (the relevant fibre bundles here are those of cingulum and corpus callosum). As in the previous example, one can see that the most accurate reconstruction is attained by means of the proposed RDG-TV method. The superiority of RDG-TV is also confirmed by the quantitative figures of Table 4.2, which summarizes the NMSE obtained by the compared algorithms for different values of K.

Chapter 5

Hybrid Blind Deconvolution using BTS

Image deconvolution refers to the process of image restoration when the original image is degraded by the action of a linear shift invariant system. When the point spread function (PSF) of the imaging system is not known *a priori*, the resulting deconvolution process is referred to as *blind* [19]. In this case, both the PSF and the original image need to be recovered directly from the observed data. Such reconstructions are especially important in many practical fields which include medical imaging [78, 79, 80, 41], astronomy [76, 75], and remote sensing [77], just to name a few.

In this chapter, we consider the problem of blind deconvolution as an inverse problem and address its solution using a novel approach. Similar to the previous chapter, the framework of the proposed solution can be considered to be composed of composite constraints arising from multiple domains. Specifically, in the present case, the problem formulation involves minimization of an objective function that involves terms both from the spatial domain as well as the frequency domain. Because of the presence of composite constraints in the optimization problem, the solution becomes computationally very complex. In order to overcome the problem, we use the tools of BTS and demonstrate how the concept of BTS can decouple the problems defined on the two domains and enable its efficient solution.

5.1 Image Reconstruction by Inverse Filtering

In many imaging modalities, the observed image g is assumed to be related to the original image f by a convolutional model given by

$$g = h * f + \eta \tag{5.1}$$

where h denotes the PSF of the imaging system in use, $*$ stands for the operation of convolution, and η accounts for the effect of both measurement and model noises. The convolution of the original image f with the band-limited PSF h in (5.1) results in an attenuation (or even complete suppression) of some spectral components of the former. These attenuated components can be recovered through the process of inverse filtering, viz.

$$f \simeq s * g, \tag{5.2}$$

where s is referred to as an *inverse filter* or a deconvolution kernel [70, 19, 71]. It should be noted that in the case of blind deconvolution, there is always a scale and linear phase ambiguity problem and so, in general, the relation in (5.2) should be replaced by

$$s * g = \alpha f(n - n_0) \tag{5.3}$$

where α is an arbitrary scalar and n_0 designates an arbitrary spatial shift of the estimate. However, for convenience of presentation of the material, we are going to use (5.2) with the ambiguity implicitly assumed.

5.2 Hybrid Blind Deconvolution

In general, the methods of blind deconvolution approaches can be categorized into two main classes. The methods of the first class, often referred to as *a priori* blur identification methods, perform estimation of the PSF as an initial step [78, 79, 85, 20, 80]. Subsequently, the estimated PSF is used to recover the original image through non-blind deconvolution. The methods of the second class, on the other hand, estimate the PSF and the original image [21, 19, 71] concurrently. Unfortunately, while the methods of the first class can be applied only to a restricted number of deconvolution scenarios, those of the second class are typically prone to the problem of local minima, let alone the relatively high computation burden imposed by their solution.

A different approach to the problem of blind deconvolution has been proposed in [41]. The main attribute of this method consists in its ability to estimate the original image

while using only partial information about the PSF, namely its power spectrum. Indeed, in many practical settings, the power spectrum can be either known from the geometry of the imaging aperture in use [168] or estimated directly from observed data [80]. Subsequently, having the partial information available, the method of [41] recovers an optimal inverse filter [19, 71] as a minimizer of a cost functional, whose definition combines the partial information with some reasonable assumptions on the properties of the image to be recovered. It should be noted that the present method can be ascribed neither to the group of deconvolution techniques which fully recover the PSF prior to estimating the reflectivity function, nor to those which recover the PSF and the reflectivity function concurrently. Hence the method has been referred to as a "hybrid" approach.

The thesis demonstrates the importance of the splitting method to solve the optimization problem which is required to find an estimation of the inverse filter. Specifically, in the case of hybrid deconvolution, this method leads to a particularly efficient numerical scheme, which is implemented as a succession of analytically computable operations. Thus, the proposed deconvolution algorithm is particularly suited for applications where fast numerical processing is critical. It is also shown how the inverse filters designed in this way can be used to deconvolve images in a non-blind manner so as to further improve their resolution and contrast.

5.3 Problem Formulation

The images f and g in (5.2) can be manipulated as $N \times M$ matrices. The inverse filter s, on the other hand, is assumed to be of the size $L_x \times L_y$, with $L_x \ll N$ and $L_y \ll M$. Moreover, to alleviate the problem of ambiguity in the linear phase of inverse filtering [71], both L_x and L_y are set to be odd integers and the central element of s is assumed to have zero delay. In this case, the $N \times M$ discrete Fourier transform (DFT) S of s can be computed as

$$S = \mathcal{F}\{s\} := F_{N\times L_x}\, s\, F^T_{M\times L_y}, \tag{5.4}$$

where the Fourier transform matrices $F_{N\times L_x}$ and $F_{M\times L_y}$ are given by

$$F_{N\times L_x}(k,n) = \exp\left\{-\jmath\frac{2\pi}{N}\,k\,n\right\}, \text{ with} \tag{5.5}$$
$$k = 0,\ldots,N-1,\; n = -\frac{L_x-1}{2},\ldots,\frac{L_x-1}{2}$$

and

$$F_{M\times L_y}(k,m) = \exp\left\{-\jmath\frac{2\pi}{N}\,k\,m\right\},\ \text{with} \tag{5.6}$$
$$k = 0,\ldots,M-1,\ m = -\frac{L_y-1}{2},\ldots,\frac{L_y-1}{2}.$$

As mentioned before, the concept of hybrid deconvolution rests on the possibility to recover partial information of the PSF h. In particular, we assume that an estimate $|\tilde{H}|$ of the magnitude spectrum of h is available. This information can be used to constrain the inverse filter to obey $||\tilde{H}|\cdot S| \approx 1$ within the passband of the imaging system in use. Here $|\tilde{H}|$ is assumed to be an $N\times M$ real-valued array and the dot stands for element-wise product. Alternatively, this condition can be restated as the requirement that the optimal s^* should be a solution of the following optimization problem

$$s^* \in \arg\min_s \left\{\frac{1}{2}\left\|W\cdot\left(\left||\tilde{H}|\cdot\mathcal{F}\{s\}\right|-1\right)\right\|_{\mathbb{F}}^2\right\} \tag{5.7}$$

where the subscript $\mathbb{F}$ stands for the Frobenius matrix norm and W is given by

$$W = \frac{|\tilde{H}|^2}{|\tilde{H}|^2+\epsilon}, \tag{5.8}$$

with $\epsilon \ll 1$. Note that W given by (5.8) has the property of being close to 1 within the imaging passband, while rapidly converging to zero on its complement.

Needless to say, the problem (5.7) is not convex, and hence it does not admit a unique minimizer. To alleviate this problem, some additional constraints on its solution need to be imposed. To this end, we note that many practical images are either sparse in nature or admit a sparse representation in the domain of a suitably chosen linear transform. This *a priori* information can be used to regularize the minimization in (5.7). In particular, in the case when f is sparse, the optimal filter s^* should be expected to minimize the ℓ_1 norm of $g * s^*$. Alternatively, in the case of f being a bounded variation image [5], minimizing the total variation (TV) of $g * s^*$ should be used instead [21]. Accordingly, the problem of finding an optimal inverse filter s^* can be reformulated as

$$s^* \in \arg\min_s \left\{\frac{1}{2}\left\|W\cdot\left(\left||\tilde{H}|\cdot\mathcal{F}\{s\}\right|-1\right)\right\|_{\mathbb{F}}^2 + \lambda\ \rho(g*s)\right\}, \tag{5.9}$$

where $\lambda > 0$ is a regularization parameter, and $\rho(x)$ is a regularization term which is equal to either $\|x\|_1$ or $\|x\|_{\mathrm{TV}}$.

5.4 Inverse Filtering using BTS

In order to apply BTS, the minimization w.r.t. s in (5.9) is substituted by minimization w.r.t. two auxiliary variables $Z := \mathcal{F}\{s\}$ and $u := g * s$. This substitution results in the following *constrained* minimization problem

$$\begin{aligned} &\min_{s,Z,u} \left\{ \frac{1}{2} \big\| W \cdot \big(\big| |\tilde{H}| \cdot Z \big| - 1 \big) \big\|_{\mathbb{F}}^2 + \lambda \ \rho(u) \right\} \\ &\text{subject to} \quad Z = \mathcal{F}\{s\}, \ u = g * s, \end{aligned} \tag{5.10}$$

which can be rewritten in an equivalent form as

$$\begin{aligned} \min_{s,Z,u} \Big\{ &\frac{1}{2} \big\| W \cdot \big(\big| |\tilde{H}| \cdot Z \big| - 1 \big) \big\|_{\mathbb{F}}^2 + \lambda \ \rho(u) + \\ &+ \frac{\delta_Z}{2} \| Z - \mathcal{F}\{s\} - b_Z \|_{\mathbb{F}}^2 + \frac{\delta_u}{2} \| u - g * s - b_u \|_{\mathbb{F}}^2 \Big\}, \end{aligned} \tag{5.11}$$

where b_Z and b_u are the Lagrangian multipliers, and $\delta_Z > 0$ and $\delta_u > 0$ are parameters whose choice is arbitrary [31]. Then, starting from some initial estimates s^0, Z^0, u^0, b_Z^0 and b_u^0 (e.g. $Z^0 = 1$, $u^0 = g$, $s^0 = \mathcal{F}^{-1}\{Z^0\}$ and $b_Z^0 = b_u^0 = 0$), the problem (5.11) can be solved sequentially as

$$s^{t+1} = \arg\min_{s} \left\{ \frac{\delta_Z}{2} \| Z^t - \mathcal{F}\{s\} - b_Z^t \|_{\mathbb{F}}^2 + \frac{\delta_u}{2} \| u^t - g * s - b_u^t \|_{\mathbb{F}}^2 \right\} \tag{5.12}$$

$$Z^{t+1} = \arg\min_{Z} \left\{ \frac{1}{2} \big\| W \cdot \big(|\tilde{H}| \cdot |Z| - 1 \big) \big\|_{\mathbb{F}}^2 + \frac{\delta_Z}{2} \| Z - \mathcal{F}\{s^{t+1}\} - b_Z^t \|_{\mathbb{F}}^2 \right\} \tag{5.13}$$

$$u^{t+1} = \arg\min_{u} \left\{ \frac{\delta_u}{2} \| u - g * s^{t+1} - b_u^t \|_{\mathbb{F}}^2 + \lambda \ \rho(u) \right\}, \tag{5.14}$$

followed by updating the Lagrange multipliers according to

$$b_Z^{t+1} = b_Z^t + \mathcal{F}\{s^{t+1}\} - Z^{t+1} \tag{5.15}$$

$$b_u^{t+1} = b_u^t + g * s^{t+1} - u^{t+1}. \tag{5.16}$$

It should be pointed out that the minimization problems (5.12)-(5.14) admit unique global minimizers (and hence the equality signs in the equations), which can be computed *analytically* as detailed below.

By virtue of Parseval's theorem, problem (5.12) can be equivalently solved in the frequency domain w.r.t. $S = \mathcal{F}\{s\}$. Particularly, denoting by U^t, G, and B_u^t the DFTs of u^t, g, and b_u^t, respectively, problem (5.12) can be alternatively formulated w.r.t. S as given by

$$S^{t+1} = \arg\min_{S} \left\{ \frac{\delta_Z}{2} \|Z^t - S - b_Z^t\|_{\mathbb{F}}^2 + \frac{\mu}{2} \|U^t - G \cdot S - B_u^t\|_{\mathbb{F}}^2 \right\}, \tag{5.17}$$

where $\mu = \delta_u / NM$. The solution S^{t+1} is unique and can be computed in a closed form as

$$S^{t+1} = \frac{\delta_z (Z^t - b_Z^t) + \mu \, \bar{G} \, (U^t - B_u^t))}{\mu \, |G|^2 + \delta_Z}, \tag{5.18}$$

where $\bar{G}$ denotes the complex conjugate of G. Having S^{t+1} computed, its spatial-domain counterpart is computed by means of the inverse Discrete Fourier Transform.

The minimization problem (5.13) is non-convex and moreover it admits more than one local minimizer. Fortunately, a global minimizer of this problem also exists. In order to obtain the solution, let us first state the following proposition.

Proposition 5.4.1. *The solution to the problem*

$$\arg\min_{z} \frac{1}{2}(|z|a - 1)^2 + \frac{\mu}{2}|zB - C|^2 \tag{5.19}$$

where $a, \mu \in \mathbb{R}_+$, and $B, C \in \mathbb{C}$, is given by

$$z = \text{rampstep}_\alpha(\beta e^{-\jmath\theta}) \, e^{\jmath\theta} \tag{5.20}$$

where

$$\alpha = \frac{a}{a^2 + \mu|B|^2}, \quad \beta = \frac{\mu C \bar{B}}{a^2 + \mu|\bar{B}|^2} \tag{5.21}$$

while θ is the phase of the complex-valued β which is given by

$$\theta = \tan^{-1}\left(\Im\{\beta\}/\Re\{\beta\}\right), \tag{5.22}$$

and the "rampstep" function (see Fig. 5.1) is defined as

$$\text{rampstep}_\alpha(x) = \begin{cases} x + \alpha, & \text{if } x \geq 0 \\ x - \alpha, & \text{if } x < 0, \end{cases} \tag{5.23}$$

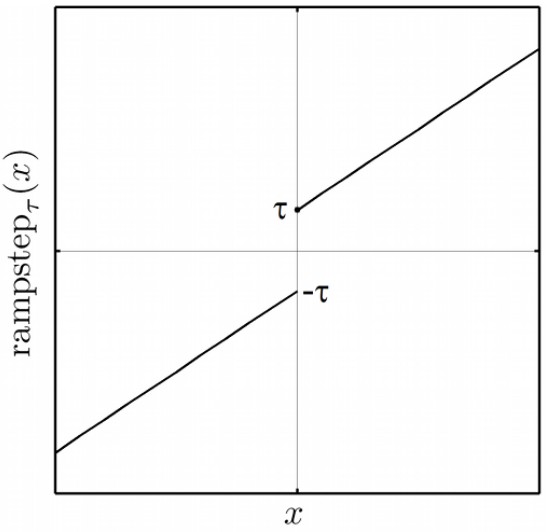


Figure 5.1: The rampstep function

The proof of proposition 5.4.1 is provided in Appendix A. Subsequently, the solution of (5.13) is obtained as follows. First, we define

$$\alpha = \frac{|\tilde{H}| \cdot W^2}{|\tilde{H}|^2 \cdot W^2 + \delta_Z}, \quad \beta = \frac{\delta_Z(\mathcal{F}\{s^{t+1}\} + b_Z^t)}{|\tilde{H}|^2 \cdot W^2 + \delta_Z}. \tag{5.24}$$

Then, the optimal solution to (5.13) is given by

$$Z^{t+1} = \text{rampstep}_\alpha(\beta e^{-\jmath\theta})\, e^{\jmath\theta}, \tag{5.25}$$

As a final step, we need to specify the solution to the problem in (5.14). In particular, the solution depends on the choice of the function ρ. For the case, when $\rho(x) = \|x\|_1$, the solution is easily computable by means of soft thresholding, as given by

$$u^{t+1} = \text{soft}_{\lambda/\delta_u}\left\{g * s^{t+1} + b_u^t\right\}, \tag{5.26}$$

On the other hand, when $\rho(x) = \|x\|_{\text{TV}}$, the solution can be obtained using, e.g., the method proposed in [143].

5.4.1 Convergence

The variable u computed in step (5.14) of the BTS algorithm represents the result of inverse filtering of the data image g with the inverse filter s. In this work, the algorithm

convergence has been controlled by monitoring the relative change in u through the course of its update, and the algorithm was terminated at the point when this change dropped below 0.1%.

5.4.2 Computational Complexity

The computational complexity of the proposed approach at each iteration can be analysed by considering the complexity associated with (5.12), (5.13) and (5.14). In particular, the complexity of (5.12) is dominated by the computation of the direct and inverse discrete Fourier transform operations whose complexity is $\mathcal{O}(NM \log_2 NM)$. On the other hand, the solution of (5.13) consists of pointwise multiplications followed by the application of the rampstep function, thus resulting in linear complexity. The complexity of (5.14) depends on the function ρ. When ρ is equal to the ℓ_1 norm, the solution is given by the soft thresholding operation which has linear complexity. On the other hand, the solution corresponding to $\rho(x) = \|x\|_{\text{TV}}$ is iterative and, for typical size of the images, can be obtained within fraction of a second in modern day digital computers.

It should be noted here that unlike the solution method provided in [41], the proposed method does not require solving system of equations to apply the Newton algorithm, and thus can be implemented with significantly improved computational complexity. Moreover, apart from the computational issues, the approach in [41] requires replacing the ℓ_1 norm by its smooth approximation to apply the smooth optimization techniques, and therefore the proposed approach can be considered to be more accurate theoretically.

5.5 Non-Blind Deconvolution

The recovered inverse filter $\tilde{s}$ can be used to estimate the PSF h. Specifically, since $\tilde{s}$ is inverse w.r.t. h, the Fourier phases of these two functions should be approximately negatives of each other. Let φ be the Fourier phase of $\tilde{S} = \mathcal{F}\{\tilde{s}\}$. Then, the Fourier transform of the PSF can be estimated as

$$\tilde{H} = |\tilde{H}|\, e^{-\jmath\varphi}, \tag{5.27}$$

which could be subsequently transformed into the spatial domain to result in an estimate $\tilde{h}$ of the PSF h.

The estimated PSF $\tilde{h}$ can be used to deconvolve the image g in a non-blind manner. Particularly, congruent with the assumption used in (5.9) is to compute an estimate $\tilde{f}$ of

the original image f according to

$$\tilde{f} = \arg\min_{f} \left\{ \frac{1}{2} \|f * \tilde{h} - g\|_{\mathbb{F}}^2 + \gamma \ \rho(f) \right\}, \tag{5.28}$$

where $\gamma > 0$ is a regularization parameter and $\rho(x)$ can be the ℓ_1 norm or TV seminorm depending on the type of image. There exist a range of numerical tools which can be employed to solve the minimization problem (5.28), and in this work we use the method of FISTA (discussed in Chapter 3) to solve the problem.

The method of hybrid blind deconvolution using BTS can be applied to a range of practical problems. In this thesis, we demonstrate the use of the technique on ultrasound images to improve its resolution and contrast.

5.6 Ultrasound Imaging

Medical ultrasound is a diagnostic technique used to visualize and examine the internal organs and soft tissues of a human body using high frequency sound waves [169, 15]. The ultrasound imaging is often preferred over other imaging modalities because it is non-invasive, inexpensive, portable, painless and does not use any ionizing radiation.

Under some reasonable assumption, the backscattered signal and the tissue reflectivity functions follow the convolution model of (5.1) [169, 20, 15, 41]. Specifically, in the case of ultrasound imaging, the acquired RF-image (g in (5.1)) is considered to be a result of the convolution of the point-spread function (PSF) of the imaging system (h in (5.1)) with the tissue reflectivity function (f in (5.1)). It should be noted that instead of working with the real valued RF images, one can work with the complex valued *in phase/quadrature* (IQ) images, which can be obtained through the process of frequency demodulation. Since the frequency demodulation is followed by anti-aliasing filter and downsampling, working with IQ images can be advantageous because of its higher signal to noise ratio and smaller size. Moreover, due to the linearity of the demodulation process, the convolution model of formation of the RF images (5.1) can also be used to describe the formation of IQ images. However, it should be noted that as opposed to the real valued RF images, the quantities in (5.1) for the IQ image are complex-valued.

The model (5.1) assumes that the PSF is spatially invariant. It should be noted that the non-uniformity of acoustic focusing, diffraction effects, dispersive attenuation and several other factors lead to the spatial variability of the PSF of ultrasound scanners [170, 41]. The smooth character of this variability, however, makes it possible to partition an ultrasound

image into a set of smaller (possibly overlapping) segments, within which the PSF can be assumed to be (approximately) spatially invariant. Yet, gaining the invariance comes with a price, as now the PSFs associated with different image segments have to be estimated (either partially or in full) as part of the image reconstruction process. Consequently, the image segments can be processed separately using the model (5.1) with corresponding local PSF, and, subsequently, the entire image is recovered by combining together the local results obtained in this process. The necessity to perform multiple deconvolutions creates a need for computationally efficient solutions which is the main objective of the present research.

5.6.1 Hybrid Deconvolution in Ultrasound Imaging

As mentioned earlier, the concept of hybrid deconvolution rests on the possibility to recover partial information on the PSF h. In the case of ultrasound imaging, the magnitude spectrum $|H|$ has been recovered using a procedure similar to that detailed in [80]. In particular, ignoring the noise term in (5.1) for the sake of simplicity, one obtains

$$G \approx H \cdot F \tag{5.29}$$

where G, H, and F are the DFTs of g, h and f, respectively. Consequently, the logarithms of their magnitude spectra bear a linear relationship as given by

$$\log|G| \approx \log|H| + \log|F| \tag{5.30}$$

Since $\log|H|$ is typically a much smoother function as compared to $\log|F|$, it can be recovered from $\log|G|$ by means of an appropriate smoothing procedure. Specifically, it has been done in [80] by subjecting $\log|G|$ to the procedure of *outlier-resistant wavelet de-noising*.

Finally, for the case of ultrasound images, one needs to select the appropriate function ρ in (5.9). In this context, it was argued in [80] that for the case when the tissue under investigation is composed of diffusive scatterers (e.g., tissue fibers, diminutive aggregates of individual cells, etc.) superimposed on a scanty structure of a few strong specular reflectors (e.g., liver arterioles, organ boundaries, etc.), the corresponding tissue reflectivity function is likely to be a *sparse* signal [171]. Such a sparse behaviour can be modelled by a Laplacian distribution, in which case the function ρ is modelled by the ℓ_1 norm [41].

5.7 Results

For the sake of convenience of referencing, the BTS-based hybrid inverse filtering method proposed in this thesis will be referred below as *sparse inverse filtering* (SIF) and the reconstruction method based on the solution of (5.28) is referred to as *sparse deconvolution* (SD).

In the experimental stage, the proposed deconvolution algorithms were compared using *in vivo* ultrasound data. To this end, a set of RF-images was recorded from adult volunteers with a VIVID3 (GE Medical ultrasound, Inc.) commercial ultrasound scanner equipped with a special data-transfer board. The recorded set of images comprised of abdominal images acquired using a linear array transducer with central frequency of 3.5 MHz. All the images were acquired with a single transmission focal point, localized approximately at the centre of the field of view. The sampling rate and resolution of data acquisition were 20 MHz and 14 bits, respectively.

The deconvolution results were similar for all the acquired images. Some typical reconstruction examples are depicted in Fig. 5.2, whose left column shows the original image of central abdomen along with its reconstructions obtained with the SIF and SD algorithms. Clearly, SD provides the results of superior quality as judged by the apparent gain in both image resolution and contrast. SIF also produces speckle pattern which is noticeably "finer" than that of the original image. However, it is lacking of high frequency components, which stems from the inability of inverse filtering to fully recover the spectral components of f which have been attenuated by the PSF.

To further support the above conclusions, the autocorrelation functions of the original and reconstructed images have been computed (as shown in the right column of Fig. 5.2). As a numerical measure of resolution, we chose the -3 dB width of the autocorrelation functions in the axial and lateral directions. The resolution gains of SIF in the axial and lateral directions were found to be 1.27 and 1.59 respectively, while those of SD were 3.44 and 2.30, respectively.

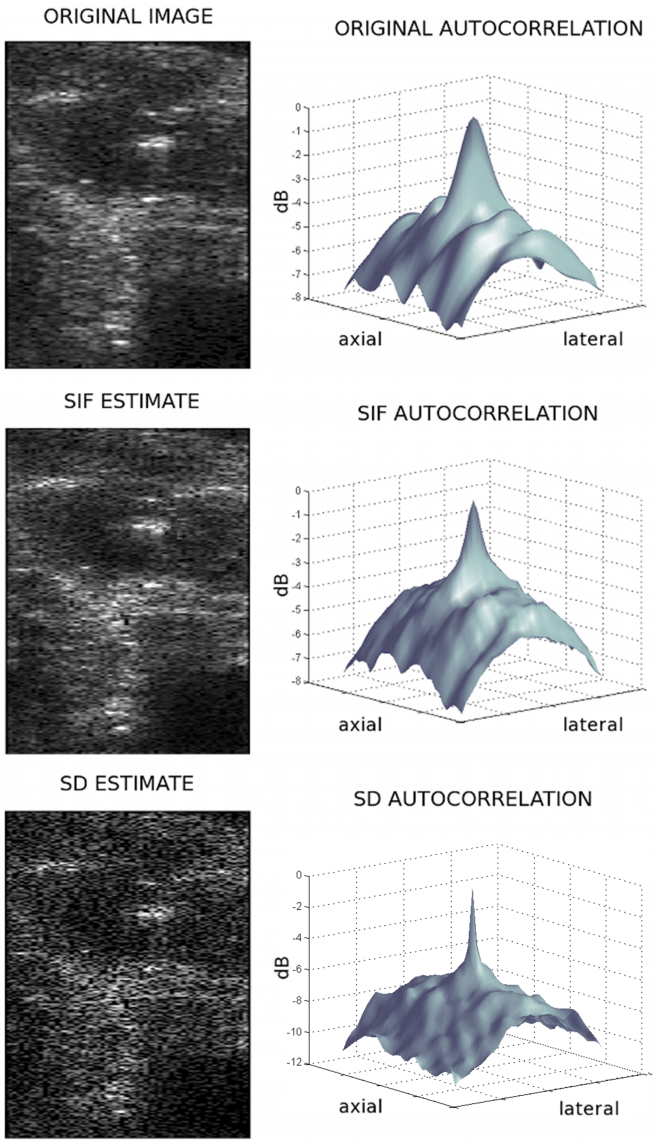


Figure 5.2: (Left column) The original ultrasound image of central abdomen and its reconstructions obtained with SIF and SD; (Right column) The autocorrelation functions of the original and reconstructed images.

Chapter 6

FCS for General Imaging

Image denoising, either as an independent application or an initial stage for subsequent processing tasks, is known to be one of the most fundamental problems in image processing. Due to possible imperfections of image acquisition and/or environmental effects, real-life images are frequently contaminated by noises. Such noises tend to obscure important details of the images, thereby reducing their overall quality. Even though the human visual system is known to be fairly resistant to the effect of noises, the latter can still constitute a serious practical problem in applications requiring automatic image processing. Accordingly, the fundamental goal of image denoising is to recover a close approximation of the original image from its noise-contaminated measurements. With the massive proliferation of digital images in the last few decades, the need for efficient denoising algorithms has been ever increasing. However, the computational complexity of the algorithms increase with the performance of the algorithm in suppressing the background noise while keeping the principal details and features of processed images as intact as possible [10, 38].

In this thesis, we introduce the formal concept of *fuzzy clustering splitting* (FCS) which supersedes a complex estimation problem by a sequence of simpler sub-problems. The goal of such a splitting is to obtain a significant reduction in computational complexity while providing reasonable quality of reconstruction. In the next section, we reiterate the main idea of FCS which has already been introduced in Chapter 1. Subsequently, we address the problem of denoising for the case of a number of key statistical distributions of contaminating noise. It should be noted that in our assumption, the noise is assumed to be independent which is a limitation of the proposed method. For example, if we have structural noise (e.g. aliasing or blocking artifact), it cannot be removed by the proposed approach. This limitation needs to be addressed in future research.

6.1 Idea of FC Splitting

As already discussed in Chapter 1, the noisy image can be considered to be a realization of a random process which can be characterized by its joint probability density function. The random process is parametrized by the underlying noise free intensities and the procedure of estimating them by considering the complete joint density function is computationally complex. Consequently, certain reasonable simplifications are required to obtain a computationally tractable estimation method.

The method of FCS performs the simplification by considering the self similarity characteristics of natural images. In particular, the pixel intensities in an image repeat themselves and this suggests that one can consider the random process as a mixture of ergodic random processes where each process is associated with those pixels having the same noise free intensity. Each such process is characterized by a single parameter which is same as the underlying noise free intensity corresponding to those pixels. Subsequently, each parameter corresponding to the associated process can be estimated separately without considering the dependencies between the processes. Accordingly, the process of FCS supersedes the complex estimation problem into two sub-problems, viz.

1. clustering of the pixels corresponding to each ergodic process, and
2. estimation of the parameter of each process based on the pixels within the corresponding cluster.

One can perform a hard strict partitioning clustering (e.g. k-means clustering) where each pixel belongs to exactly one cluster. However, in general, this approach requires a priori knowledge of the number of clusters which might not be available in practical scenarios. Moreover, the assignment of the pixels to the clusters has to be done based on noisy data and hence is prone to errors. In order to overcome the aforementioned problems, FCS suggests using a soft clustering where a cluster is defined corresponding to each pixel. Subsequently, the other pixels are assigned to the particular cluster with certain probability which can be taken into account while estimating the parameter of the process. For convenience of referencing, let us define the pixel to be estimated as the *source pixel* and the other members within the cluster as the *target pixels*. Accordingly, our problem is to estimate the original intensity of the source pixel based on noisy intensities of the target pixels.

Although the FCS has not been proposed formally before, the underlying principles have been used for decades in several denoising procedures. Thus, for example, a particular type

of clustering can be obtained by assuming that the neighbouring pixels of the source pixel belong to the same ergodic process. A specific instance of such an approach is the linear space invariant filter discussed in Section 3.4 of Chapter 3. In this case, the target pixels are assigned to the cluster with probabilities which are based on their distance and orientation with respect to the source pixel. Once the fuzzy clustering is performed, the estimate of the source pixel is obtained using a weighted averaging scheme where the weights corresponding to the target pixels are dependent on their membership probability within the cluster. Unfortunately, such a naive approach suffers from severe blurring artifacts at the edges or areas of fine structures where the assumption of neighbourhood similarity fails. The problem has been alleviated using range [38] and bilateral filters [56, 57, 39] which are also discussed in Chapter 3. In these filters as well, the pixel intensity is estimated as a weighted average of the neighbouring pixels. However, in contrary to the linear space invariant filters, the weights are assigned to the target pixels based on their similarities with the source pixel. Because of the fact that the source pixels are estimated based on the intensities of their neighbourhood pixels, the filters discussed above are often referred to as neighbourhood filters.

The range filters and bilateral filters can remove mild to medium level of noise. However, the accuracy of the fuzzy clustering decreases with increase in the level of noise. Yet even more accurate reconstruction results are obtainable with *non-local means* (NLM) filtering, which has been shown to outperform a wide range of alternative denoising techniques for the case of additive white Gaussian (AWG) noise contamination [9, 10]. The non-local means filter uses the a priori information that the pixel intensities are not the only attributes that repeat themselves, the patches within an image tend to do the same as well. Accordingly, the weights can be computed based on the similarity of the patches around the source and the target pixels as opposed to the similarity between the pixels themselves. This method is expected to be more robust as compared to finding the similarity based on a single pair of pixels because noise is less likely to affect equally all the pixels within a patch. Hence, in this thesis, we compute the fuzzy clustering probability based on the similarity between patches.

The original version of the non-local means algorithm [9, 10] has been designed for removal of additive white Gaussian noise. Subsequently, several theoretical frameworks have been developed aiming to incorporate more general noise statistics [68, 7, 69]. Unfortunately, none of these approaches could be regarded as general, as it has been found that the NLM method which is optimal for a particular type of noise is likely to produce inconsistent denoising results for a different type of noise, as explained below.

Our contribution in this chapter is two-fold. First, it proposes the formal framework of the fuzzy clustering splitting. Second, we propose an original approach to perform the

clustering in a unifying way which is independent of the type of the measurement noise. Specifically, as will be discussed below, most of the approaches of performing the clustering is highly dependent on the nature of the noise to be rejected, and hence, is not applicable when the assumptions are not met. This fact will be addressed in this chapter with the proposal of a novel methodology.

Before proceeding further, let us briefly revisit the concept of neighbourhood filtering which has already been discussed in Chapter 3.

6.2 Neighbourhood Filter

Let X_s and Y_s denote the intensities of an original image X and its observed noisy counterpart Y, respectively, corresponding to the spatial location $s \in \Omega \equiv \{0, 1, \ldots, N-1\}$, where N denotes the total number of pixels in the image. In such a case, Y can be considered to be a realization of a mixture of ergodic random processes with each process being dependent on the underlying noise free intensities. In the case when X_s represents the mean value of Y_s, an estimate of X_s can be obtained by averaging $\{Y_t\}_{t \in J_s}$, where J_s denotes the cluster containing the indices whose associated intensities are distributed identically to Y_s. As mentioned before, since one does not have access to J_s for each s, it makes sense to replace the averaging by a weighted averaging which will reflect the effect of fuzzy clustering. Formally, in such a case, the estimate of X_s is computed according to

$$\hat{X}_s = \frac{1}{C_s} \sum_{t \in \Omega} w_{s,t} Y_t, \text{ with } C_s = \sum_{t \in \Omega} w_{s,t}, \tag{6.1}$$

where the *weights* $w_{s,t} \geq 0$ quantify the contribution of target pixel intensities $Y_t,\ t \in \Omega$, to denoising of the source intensity at $s \in \Omega$. In general, the weight $w_{s,t}$ should reflect the probability of fuzzy clustering which, in turn, depends on the degree of similarity between the image intensities at s and t. As discussed in Chapter 3, in the case of range filtering [38], the weights are defined as

$$w_{s,t} = \exp\left(-\frac{|Y_s - Y_t|^2}{h^2}\right). \tag{6.2}$$

From (6.2), one can see that the weight $w_{s,t}$ increases when the difference $|Y_s - Y_t|$ becomes smaller and vice versa. The parameter $h > 0$ controls the amount of filtering, with higher values of h producing more profound filtering effect. Consequently, h should be set to maintain an optimal balance between the removal of noise and preservation of fine image details.

In its original version, NLM filtering [9] uses a definition of weights $w_{s,t}$ similar to that in (6.2), while replacing the absolute difference between the intensities Y_s and Y_t by a β_k-weighted Euclidean distance between the image patches centred around Y_s and Y_t, respectively. Formally,

$$w_{s,t} = \exp\left(-\frac{1}{h^2}\sum_{k\in I}\beta_k\,|Y_{s-k} - Y_{t-k}|^2\right), \tag{6.3}$$

where I is the set of pixel indices which represent a symmetric neighbourhood of the centre of image coordinates. (Thus, for example, with s corresponding to two spatial coordinates s_1 and s_2, i.e. $s = (s_1, s_2)$, I can be defined as $I = \{(s_1, s_2) \in \mathbb{Z}^2 : |s_1| \leq L_1, |s_2| \leq L_2\}$ for some positive integers L_1 and L_2.) The "fine-tuning" parameters $\{\beta_k\}_{k\in\Omega}$ in (6.3) are intended to weigh the domain of summation and they are usually chosen to satisfy $\sum_{k\in\Omega}\beta_k = 1$, while the parameter $h > 0$, as before, controls the overall amount of smoothing imposed by the filter. For the case when $\beta_k/h = 1$, $\forall k$, the exponent in (6.3) transforms into the negative Euclidean distance between the intensities of the two image patches centred at pixels s and t, respectively. Moreover, it should be noted that the expression of the weights in (6.3) has been shown to be justified in the case of additive white Gaussian (AWG) noise [68, 7].

To facilitate our considerations, we note that the NLM weights in (6.3) can be expressed in an alternative way as given by [7, 69]

$$w_{s,t} = \prod_{k\in I}(\mathrm{SNL}_{s,t,k})^{\frac{\beta_k}{\alpha}}, \tag{6.4}$$

with $\mathrm{SNL}_{s,t,k}$ being a *Gaussian similarity measure* (SM) defined as

$$\mathrm{SNL}_{s,t,k} = \exp\left(-|Y_{s-k} - Y_{t-k}|^2/4h^2\right), \tag{6.5}$$

and $\alpha = 1/4$. A similarity measure reflects the similarity between a pair of pixels. This measure, in turn, is used to perform the fuzzy clustering as well as to compute the weights in (6.4). In general, the parameter $\alpha > 0$ controls the dependence between the pixels in each of the compared image patches, and is kept as a tuning parameter in the algorithm [7]. It is important to note that the Gaussian SM is always bounded between 0 and 1, and it has been shown to be justified in the case of additive white Gaussian (AWG) noise contamination (see [68, 7] for more details). Moreover, it should be noted that two SMs with exponents differing by a factor can produce the same weight (by adjusting α) and, therefore, from the algorithmic point of view, they can be considered to be equivalent. This fact will be utilized further in this chapter.

As noted above, the Gaussian SM in (6.5) along with its associated weights (6.3) are optimal for NLM filtering of images contaminated by AWG noise [68, 7]. It is known, however, that using this SM is likely to produce suboptimal denoising results for different types of noise contamination. For example, in the case of a multiplicative noise (where the noise intensity depends on that of the underlying original image), the weights (6.3) tend to overly blur the image at low-intensity areas, while letting excessive levels of noise at high-intensity areas. This fact advocates the need for adjusting the formulation of SM to the statistical properties of noise. In the next section, we overview some existing SM formulations derived to this end, followed by a discussion of their main drawbacks.

6.3 Similarity Measures

The success of NLM filtering in application to denoising of images contaminated by AWG noise has motivated researchers to extend the algorithm to other types of measurement noises. However, some earlier works undertaken in this direction have been based either on a direct application of the Gaussian SM to different types of noise (such as, e.g., Rician noise [172]) or on *ad hoc* adaptations of the same SM to the specific noise statistics in question ([66, 173]). Unfortunately, neither of the aforementioned approaches provides a theoretical framework which could be used for the derivation of optimal, noise-specific SMs in a unified and consistent manner. Addressing this problem has become the main objective of a number of subsequent studies [68, 7, 67, 69]. Before discussing some principal shortcomings of the existing methods of NLM filtering, it is instructive to point out a number of desirable properties of SMs.

P1: The similarity measure $\mathrm{SNL}_{s,t,k}$ should attain its maximum value when the two observed intensities Y_{s-k} and Y_{t-k} are equal. Moreover, the SM should decrease monotonically as a function of $|Y_{s-k} - Y_{t-k}|$.

P2: The SM should be symmetric, which implies

$$\mathrm{SNL}_{s,t,k} = \mathrm{SNL}_{t,s,k}. \tag{6.6}$$

P3: The SM should be bounded. If that is not the case then the weight assigned to a certain pixel might dominate over the other weights, thereby preventing the algorithms from effectively removing the noise.

Combining all the above conditions results in

$$\mathrm{SNL}_{s,t,k} = \mathrm{SNL}_{t,s,k} \leq \mathrm{SNL}_{s,s,k} = \mathrm{SNL}_{t,t,k} < \infty \tag{6.7}$$

Though very intuitive, the above properties are not always met by SMs proposed in the literature. Thus, for instance, the work in [68] has proposed an elegant formulation of the SM and NLM filtering of [10, 9] based on a Bayesian estimation framework. Unfortunately, this formulation can be shown to result in SMs disobeying Properties P1-P3 for asymmetric probability distributions of measurement noises.

As was mentioned above, several works have been devoted to developing a unified approach to the formulation of SMs for arbitrary noise distributions. Among those, the most relevant to the present work are the methods reported in [7, 69]. Specifically, in [7], the SM has been set to be equal to the posterior probability of the event $X_{s-k} = X_{t-k}$ conditioned on the observations of Y_{s-k} and Y_{t-k}. Formally,

$$\mathrm{SNL}_{s,t,k} = P(X_{s-k} = X_{t-k} \,|\, Y_{s-k}, Y_{t-k}). \tag{6.8}$$

In this case, assuming $\beta_k/\alpha = 1,\ \forall k \in I$, the expression for the NLM weights $w_{s,t}$ in (6.4) becomes

$$w_{s,t} = \prod_{k \in I} P(X_{s-k} = X_{t-k} \,|\, Y_{s-k}, Y_{t-k}). \tag{6.9}$$

Under the assumption of statistical independence of the intensities of the original image X, the weights in (6.9) can be interpreted as the posterior probability of the image patches $\{X_{s-k}\}_{k \in I}$ and $\{X_{t-k}\}_{k \in I}$ to have identical intensities, conditioned on the observations of their noisy counterparts $\{Y_{s-k}\}_{k \in I}$ and $\{Y_{t-k}\}_{k \in I}$, respectively. Subsequently, using Bayes' theorem and assuming the prior probability of Y to be uniform, one obtains [7]

$$\mathrm{SNL}_{s,t,k}(y_1, y_2) = \int p_{Y_{s-k} \,|\, X_{s-k}}(y_1 \,|\, x)\, p_{Y_{t-k} \,|\, X_{t-k}}(y_2 \,|\, x)\, p_X(x)\, dx, \tag{6.10}$$

where $p_X(x)$ denotes the prior distribution of the intensities of the original image X. (Note that the same definition of SM was also employed in [174].) In the absence of *a priori* information on $p_X(x)$, the latter is normally assumed to obey an improper uniform or Jeffrey's prior model. Specifically, the uniform priors have been successfully used in a number of works [7, 69], and for this reason, we will adopt this statistical assumption here as well.

It should be noted that, in the case of continuous random variables, the probability $P(X_{s-k} = X_{t-k} \,|\, Y_{s-k}, Y_{t-k})$ is equal to zero almost surely [175, p. 111], which will yield

zero value of $\mathrm{SNL}_{s,t,k}(y_1, y_2)$ in (6.10) for any y_1 and y_2. A possible way to get around of this theoretical inconsistency is to introduce an auxiliary random variable U equal to the difference of X_{s-k} and X_{t-k}, i.e. $U \triangleq X_{s-k} - X_{t-k}$, and then analyze the conditional probability density $p_{U\,|\,Y_{s-k},Y_{t-k}}$. In this case, the SM can be set to be equal to the value of $p_{U\,|\,Y_{s-k},Y_{t-k}}$ taken at zero [69]. Formally,

$$\begin{aligned}\mathrm{SNL}_{s,t,k}(y_1, y_2) &= p_{U\,|\,Y_{s-k},Y_{t-k}}(0\,|\,y_1, y_2) \\ &= \int p_{X_{s-k}\,|\,Y_{s-k},Y_{t-k}}(x\,|\,y_1, y_2)\, p_{X_{t-k}\,|\,Y_{s-k},Y_{t-k}}(x\,|\,y_1, y_2)\, dx.\end{aligned} \tag{6.11}$$

Further, under the simplifying assumption of

$$p_{X_{z_1-k}\,|\,Y_{z_1-k},Y_{z_2-k}} = p_{X_{z_1-k}\,|\,Y_{z_1-k}}, \quad z_1, z_2 \in \Omega,\ z_1 \neq z_2, \tag{6.12}$$

and assuming the prior probabilities p_X and p_Y to be uniform, the SM in (6.11) can be further simplified to become

$$\mathrm{SNL}_{s,t,k}(y_1, y_2) = \int p_{Y_{s-k}\,|\,X_{s-k}}(y_1\,|\,x)\, p_{Y_{t-k}\,|\,X_{t-k}}(y_2\,|\,x)\, dx. \tag{6.13}$$

It turns out that the above formulation of the SM works particularly well in the case of additive noises. It was shown in [69], however, that in the case of multiplicative noises, the SMs resulting from (6.13) may no longer possess Properties P1-P3. As a solution to this problem, it was then suggested to convert multiplicative noises into additive noises by means of the logarithmic transformation. In this case, the SM in (6.13) can be shown to transform into

$$\begin{aligned}&\mathrm{SNL}_{s,t,k}(y_1, y_2) \\ &= \int p_{\log(Y_{s-k})\,|\,\log(X_{s-k})}(\log(y_1)\,|\,x)\, p_{\log(Y_{t-k})\,|\,\log(X_{t-k})}(\log(y_2)\,|\,x)\, dx.\end{aligned} \tag{6.14}$$

Unfortunately, despite their strong theoretical foundations, the methods of formulation of SMs introduced in [7] and [69] are not guaranteed to result in SMs that obey Properties P1-P3 for the case of more general noise models, which can be classified as neither additive nor multiplicative. Such cases, therefore, require a different formulation of SMs, a possible instance of which is discussed in the next section of the chapter.

6.4 Correlative Similarity Measure

6.4.1 Auxiliary Variables

As mentioned in [175, p. 111], the concept of equality between two random variables is not well defined. Accordingly, to avoid possible theoretical confusions, we follow [175, p. 111] and introduce two auxiliary random variables

$$U \triangleq X_{s-k} - X_{t-k} \quad \text{and} \quad V \triangleq X_{s-k}/X_{t-k}. \tag{6.15}$$

Using these variables, one can define the *subtractive similarity measure* (SSM) as the value of $p_{U\,|\,Y_{s-k},Y_{t-k}}(u\,|\,y_1,y_2)$ at $u=0$, i.e.,

$$\mathrm{SSM}_{s,t,k}(y_1,y_2) = p_{U\,|\,Y_{s-k},Y_{t-k}}(0\,|\,y_1,y_2), \tag{6.16}$$

as well as the *rational similarity measure* (RSM) as the value of $p_{V\,|\,Y_{s-k},Y_{t-k}}(v\,|\,y_1,y_2)$ at $v=1$, i.e.,

$$\mathrm{RSM}_{s,t,k}(y_1,y_2) = p_{V\,|\,Y_{s-k},Y_{t-k}}(1\,|\,y_1,y_2). \tag{6.17}$$

It should be pointed out that the SSM is identical to that given by (6.13) [7], while the RSM can be shown to be equal to

$$p_{\log(X_{s-k})-\log(X_{t-k})\,|\,\log(Y_{s-k}),\log(Y_{t-k})}(0\,|\,\log(y_1),\log(y_2)) \tag{6.18}$$

for the case of multiplicative noises. Moreover, using the simplifying assumption in (6.12) and assuming the prior probabilities p_X and p_Y to be uniform, the RSM can be shown to be equal to

$$\mathrm{RSM}_{s,t,k}(y_1,y_2) = \int x\, p_{Y_{s-k}\,|\,X_{s-k}}(y_1\mid x)\, p_{Y_{t-k}|X_{t-k}}(y_2\mid x)\, dx. \tag{6.19}$$

It should be noted that the RSM in (6.19) is no longer equal to (6.14) which has been introduced in [69] for the case of multiplicative noises. This is because the SM in (6.14) was derived under the assumption of uniformity of $p_{\log(X)}$ and $p_{\log(Y)}$, while that in (6.19) was derived assuming p_X and p_Y to be uniform. In what follows, it will be shown that the above assumptions may lead to substantial differences in the resulting SMs.

6.4.2 Correlative SSM

In general, neither the SSM nor the RSM are guaranteed to produce SMs satisfying Properties P1-P3. To alleviate this problem, we first note that the SSM can be interpreted as the standard $\mathbb{L}_2$ inner product between the *likelihood functions* $L_{y_1}(x) \triangleq p_{Y_1 \mid X}(y_1 \mid x)$ and $L_{y_2}(x) \triangleq p_{Y_2 \mid X}(y_2 \mid x)$, *viz.*

$$\mathrm{SSM}_{s,t,k}(y_1, y_2) \equiv \langle L_{y_1}, L_{y_2} \rangle_{\mathbb{L}_2} = \int L_{y_1}(x)\, L_{y_2}(x)\, dx. \tag{6.20}$$

Although the inner product in (6.20) is symmetric (thereby resulting in a symmetric SM), it is neither guaranteed to be bounded nor to attain its maximum value at $y_1 = y_2$. It turns out, however, that the latter two properties can be assured through a simple normalization, as given by

$$\mathrm{C}^S_{s,t,k}(y_1, y_2) = \frac{\langle L_{y_1}, L_{y_2} \rangle_{\mathbb{L}_2}}{\|L_{y_1}\|_{\mathbb{L}_2} \|L_{y_2}\|_{\mathbb{L}_2}}, \tag{6.21}$$

where $\|L_{y_i}\|_{\mathbb{L}_2} = \sqrt{\langle L_{y_i}, L_{y_i} \rangle_{\mathbb{L}_2}}$, $i \in \{1, 2\}$ and the superscript S stands for "subtractive". This new SM will be referred below as the *correlative SSM* (CSSM) (on account of its apparent similarity with the correlation coefficient used in statistical literature). It is interesting to observe that C^S can be also viewed as the inner product between two functions lying on the unit sphere in $\mathbb{L}_2(\mathbb{R})$. Hence, C^S has the interpretation of the cosine of the angle between the two functions [176]. Moreover, from the definition of C^S, the following observations can be immediately made.

1. The Cauchy-Schwartz inequality implies that

   $$C^S_{s,t,k}(y_1, y_2) \leq 1 \tag{6.22}$$

   with the equality holding if and only if $L_{y_1}(x) = L_{y_2}(x)$, which in turn happens when $y_1 = y_2$. Hence C^S is maximized when $y_1 = y_2$, as required by Property P1.

2. Similar to the inner product, $C^S_{s,t,k}(y_1, y_2)$ is symmetric, i.e. $C^S_{s,t,k}(y_1, y_2) = C^S_{s,t,k}(y_2, y_1)$, as required by Property P2.

3. From (6.22), one can also conclude that $C^S_{s,t,k}(y_1, y_2)$ is bounded, as required by Property P3.

6.4.3 Correlative RSM

The same methodology used for deriving the CSSM can be applied to the case of the RSM. To this end, we first notice that the RSM can be redefined in terms of the likelihood functions $L_{y_1}(x)$ and $L_{y_1}(x)$ as

$$\mathrm{RSM}_{s,t,k}(y_1, y_2) = \int x\, L_{y_1}(x)\, L_{y_2}(x)\, dx. \tag{6.23}$$

Similarly to (6.20), the integral in (6.23) can be interpreted as a *weighted* inner product $\langle f, g\rangle_{\mathbb{S}} = \int f(x)\, g(x)\, x\, dx$, where $x\, dx$ can be viewed as a "modified" integration measure in some real Hilbert space $\mathbb{S}$. Using this notation, the RSM can be written as

$$\mathrm{RSM}_{s,t,k}(y_1, y_2) \equiv \langle L_{y_1}, L_{y_2}\rangle_{\mathbb{S}}. \tag{6.24}$$

Subsequently, after a normalization analogous to (6.21), $\mathrm{RSM}_{s,t,k}(y_1, y_2)$ transforms into the *correlative RSM* (CRSM) given by

$$C^R_{s,t,k}(y_1, y_2) = \frac{\langle L_{y_1}, L_{y_2}\rangle_{\mathbb{S}}}{\|L_{y_1}\|_{\mathbb{S}}\|L_{y_2}\|_{\mathbb{S}}}, \tag{6.25}$$

where $\|L_{y_i}\|_{\mathbb{S}} = \sqrt{\langle L_{y_i}, L_{y_i}\rangle_{\mathbb{S}}}$, $i \in \{1, 2\}$ and the superscript R standing for "rational".

Once again, similar to the case of the CSSM, one can observe the following facts.

1. When $X > 0$, one can use the Cauchy-Schwartz inequality on the functions $\sqrt{x}L_{y_1}(x)$ and $\sqrt{x}L_{y_2}(x)$ to show that
$$C^R_{s,t,k} \leq 1. \tag{6.26}$$
with the equality attained if and only if $y_1 = y_2$. It should be emphasized, however, that the above property holds only for $X > 0$, while leading to undesirable results in the cases when X can assume negative values.

2. C^R is a symmetric SM, i.e. $C^R_{s,t,k}(y_1, y_2) = C^R_{s,t,k}(y_2, y_1)$.

3. From (6.26), one can conclude that C^R is bounded and the maximum value is attained when its arguments are equal.

In this thesis, we are going to use the formulations of CSSM and CRSM in order to obtain the fuzzy clustering splitting. In particular, we are going to obtain the closed form expressions for different types of practical noises and compare them to some existing approaches. However, before doing that, let us discuss the second stage of the splitting procedure, namely estimating the source intensity once the fuzzy cluster is obtained.

6.5 Denoising Formulation

Once the fuzzy clustering corresponding to each source pixel is obtained, the next part of the problem is to use the target pixels to obtain an estimate of the corresponding source pixel. In this context, it is worthwhile to note that the linear weighting in (6.1) is not the only way in which the weights $w_{s,t}$ can be combined with the noisy samples Y_t to yield an estimate $\hat{X}_s$. In fact it can lead to suboptimal result if the parameter of interest does not correspond to the mean of the ergodic process. Consequently, the denoising strategy should consider the noise statistics into account.

With a strict partitioning clustering, one can use a maximum likelihood (ML) approach in order to estimate the source pixel. However, such an approach is not recommended in the present case where the pixels belong to a particular cluster with certain probability. A more general approach has been suggested in[7, 69] based on the *weighted maximum likelihood* (WML) framework of [177]. According to this approach, an estimated value $\hat{X}_s$ of X_s can be computed as given by

$$\hat{X}_s = \arg\max_{X_s} \sum_{t \in \Omega} w_{s,t} \log\, p(Y_t \,|\, X_s), \tag{6.27}$$

where $p(Y_t \,|\, X_s)$ denotes the conditional probability density of Y_t given X_s. Such an approach uses the principles of ML estimation while incorporating the uncertainty presented by the fuzzy clustering.

In this thesis, the WML-based estimation of (6.27) is employed to estimate the source pixel. However, the WML based estimation scheme is not computationally tractable for all types of measurement noises. In such cases, alternative solutions will be proposed.

6.6 Closed Form Solutions

In this section, we will derive the CSSM and CRSM for some key statistical distribution of noise prevalent in a number of important imaging modalities.

6.6.1 Additive White Gaussian Noise

In the case of AWG noise contamination, using the SSM can be shown to lead to the same definition of the NLM weights as was previously proposed in [9]. Specifically, following [7],

one obtains

$$\mathrm{SSM}_{s,t,k} = \frac{1}{\sqrt{4\pi\sigma^2}} \exp\left(-\frac{(y_1 - y_2)^2}{4\sigma^2}\right). \tag{6.28}$$

Subsequently, using the fact that, at the case at hand, $\|L_{y_1}\|_{\mathbb{L}_2} = \|L_{y_2}\|_{\mathbb{L}_2} = (4\pi\sigma^2)^{-1/4}$, the CSSM can be defined as

$$C^S_{s,t,k}(y_1, y_2) = \exp\left(-\frac{(y_1 - y_2)^2}{4\sigma^2}\right) \tag{6.29}$$

Hence using a suitable value of α in (6.4), namely $\alpha = 1/4$, one obtains the same weights as it has been proposed in [9] for denoising of image contaminated by AWG noise. It is interesting to note that the same weights have been derived in [68, 7] based on different statistical considerations and assumptions.

To derive the CRSM $C^R_{s,t,k}$, the inner product $\langle L_{y_1}, L_{y_2}\rangle_{\mathbb{S}}$ needs to be computed first, which can be shown to be equal to

$$\langle L_{y_1}, L_{y_2}\rangle_{\mathbb{S}} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(y_1 - y_2)^2}{4\sigma^2}\right) \frac{y_1 + y_2}{2}. \tag{6.30}$$

Subsequently, using the fact that $\|L_{y_i}\|_{\mathbb{S}} = y_i/2\pi\sigma^2$, $i \in \{1, 2\}$, the resulting CRSM is defined by

$$C^R_{s,t,k}(y_1, y_2) \equiv \left(\sqrt{\frac{y_1}{y_2}} + \sqrt{\frac{y_2}{y_1}}\right) \exp\left(-\frac{(y_1 - y_2)^2}{4\sigma^2}\right). \tag{6.31}$$

Since a Gaussian random variable can assume negative values, the norms $\|L_{y_1}\|_{\mathbb{S}}$ and $\|L_{y_2}\|_{\mathbb{S}}$ used in the derivation of (6.31) may be negative, which is a rather unnatural result. Moreover, even when y_1 and y_2 are constrained to be strictly positive, the Gaussian CRSM may become unbounded. For this reason, $C^R_{s,t,k}(y_1, y_2)$ should not be used as a SM for NLM denoising of images contaminated by AWG noise.

6.6.2 Multiplicative Noise

In the case of multiplicative noises, the image formation model is given by

$$Y_k = X_k \cdot V_k, \quad k \in \Omega, \tag{6.32}$$

where X_k are the noise-free intensities, Y_k are their noisy counterparts, and V_k are independent and identically distributed random variables following some probability distribution.

Multiplicative noises (a particularly important example of which is speckle noise) are ubiquitous throughout many important imaging system, among which are synthetic aperture radar, sonar, ultrasound, and laser imaging [178, 179, 180, 181, 14, 182, 183, 30, 69, 47, 184, 185, 186], where the image formation is based on coherent interference of either electromagnetic or acoustic waves. Consequently, a spectrum of relevant denoising methods has been proposed in the literature [47, 184, 185, 186, 30, 69]. In the context of NLM filtering, the removal of multiplicative noise has been addressed in [69].

In our derivation of the CSSM and CRSM for multiplicative noise models, we will use the fact that, in the case of (6.32), the conditional probability density $p_{Y_k \,|\, X_k}(y \,|\, x)$ is given by

$$p_{Y_k \,|\, X_k}(y \,|\, x) = \frac{1}{|x|} \, p_{V_k}\left(\frac{y}{x}\right), \tag{6.33}$$

with $p_{V_k}(v)$ being the probability density function (pdf) of the noise V_k. Moreover, it is commonly assumed that both $p_{X_k}(u) = 0$ and $p_{V_k}(u) = 0$ for $u < 0$, and therefore $Y_k > 0$ almost surely [69].

Multiplicative Rayleigh Noise

In the context of coherent imaging system, the *fully developed speckle noise* is often modelled by a multiplicative noise model with V_k obeying a Rayleigh distribution [187, 188, 189, 182, 30, 183], whose pdf is given by

$$p_{V_k}(v) = \begin{cases} \frac{v}{\sigma^2} \exp\left(-\frac{v^2}{2\sigma^2}\right), & v \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.34}$$

where $\sigma > 0$ is a parameter of the distribution. Consequently, (6.33) suggests that the conditional pdf $p_{Y_k \,|\, X_k}(y \,|\, x)$ is given by

$$p_{Y_k | X_k}(y \mid x) = \begin{cases} \frac{y}{x^2\sigma^2} \exp\left(-\frac{y^2}{2x^2\sigma^2}\right), & y \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{6.35}$$

Next, we derive the expressions for CSSM and CRSM for the case of multiplicative Rayleigh noise.

Proposition 6.6.1. *In the case of multiplicative Rayleigh noise, the CSSM $C^S_{s,t,k}$ and CRSM $C^R_{s,t,k}$ are equal to*

$$C^S_{s,t,k}(y_1, y_2) = \left(\frac{y_1 y_2}{y_1^2 + y_2^2}\right)^{3/2} = \left(\frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{-3/2}, \tag{6.36}$$

and

$$C^R_{s,t,k}(y_1, y_2) = \frac{y_1 y_2}{y_1^2 + y_2^2} = \left(\frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{-1}. \tag{6.37}$$

respectively.

The derivations of C^S and C^R for multiplicative Rayleigh noise are provided in Appendix B.1. It should be noted that, the expressions are presented up to constant multipliers, which are irrelevant due to the normalization of NLM weights. Moreover, when $C^S_{s,t,k}$ and $C^R_{s,t,k}$ are used for computing the NLM weights $w_{s,t}$ according to (6.4) (with α being used as a user-defined "tuning" parameter), one has

$$w_{s,t} = \prod_{k \in I} (C^S_{s,t,k})^{\beta_k/\alpha_S} = \prod_{k \in I} (C^R_{s,t,k})^{\beta_k/\alpha_R} \tag{6.38}$$

for any values of α_S and α_R which satisfy $2\,\alpha_S = 3\,\alpha_R$. In other words, as long as the NLM weights are concerned, the "subtractive" and "multiplicative" formulations can be transformed into one another through appropriate choice of parameter α in (6.4). Consequently, the SMs $C^S_{s,t,k}$ and $C^R_{s,t,k}$ in (6.36) and (6.37) can be regarded as *equivalent*. For a fixed value of α, however, $C^S_{s,t,k}$ provides better contrast than $C^R_{s,t,k}$ for discriminating dissimilar values of Y, while producing poorer contrast for identifying close values of Y.

It is interesting to note that the SM proposed in [69] is given by

$$\mathrm{SNL}_{s,t,k}(y_1, y_2) = \left(\frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{-2}, \tag{6.39}$$

which is equivalent to SMs (6.36) and (6.37) in the sense of (6.38). The same study shows that, for the given type of measurement noise, the WML-optimal formulation (6.27) of NLM filtering is given by

$$\hat{X}_s = \sqrt{\frac{1}{2\sigma^2} \frac{\sum_t w_{s,t} Y_t^2}{\sum_t w_{s,t}}}. \tag{6.40}$$

Accordingly, the above formula was used in the experimental part of the present chapter as well.

Multiplicative Gamma Noise

A different model of speckle noise contamination as used in [69, 47, 30, 190, 191] suggests to describe the multiplicative noise V_k in (6.32) as a random variable obeying Gamma

distribution. The pdf of a Gamma distributed random variable is known to have the form of

$$p_V(v) = \begin{cases} \frac{m^m}{\Gamma(m)} v^{m-1} \exp(-mv), & v \geq 0 \\ 0, & \text{otherwise,} \end{cases} \tag{6.41}$$

where Γ is the Gamma function and $m \geq 1$ is a parameter of the distribution. The conditional density $p_{Y_k|X_k}(y \mid x)$ in such a case is given by

$$p_{Y_k|X_k}(y \mid x) = \frac{m^m}{\Gamma(m)} \frac{y^{m-1}}{x^m} \exp\left(-\frac{m\,y}{x}\right), \quad y \geq 0. \tag{6.42}$$

Proposition 6.6.2. *In the case of multiplicative Gamma noise contamination, the CSSM and CRSM have the following closed form expressions*

$$C^S_{s,t,k}(y_1, y_2) = \frac{(y_1 y_2)^{m-\frac{1}{2}}}{(y_1 + y_2)^{2m-1}} = \left(2 + \frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{-(m-\frac{1}{2})} \tag{6.43}$$

and

$$C^R_{s,t,k}(y_1, y_2) = \frac{(y_1 y_2)^{m-1}}{(y_1 + y_2)^{2m-2}} = \left(2 + \frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{-(m-1)}. \tag{6.44}$$

Once again, the above expressions are provided up to constant multipliers, which have no influence on NLM weights due to normalization. The derivations of the similarity measures are provided in Appendix B.2.

It should also be noted that the SMs in (6.43) and (6.44) can be regarded to be equivalent in the sense of (6.38), as their respective NLM weights can be transformed one into another through appropriate choice of parameters α in (6.4). Moreover, these SMs also appear to be equivalent to the SM derived in [69] which, after ignoring the constant terms, is given by

$$\mathrm{SNL}_{s,t,k}(y_1, y_2) = \left(2 + \frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{-m}. \tag{6.45}$$

It was also demonstrated in [69] that, for the noise distribution at hand, the WML-optimal NLM weighing scheme is given by (6.1), which is used in the present study as well.

6.6.3 Poisson Noise

The "Poissonian" degradation of digital images is typical for imaging modalities, whose data acquisition mechanism exploits the notion of even counts. Among such modalities

are, for example, positron emission tomography (PET) [192] and single positron emission computer tomography (SPECT) [193], optical imaging [194, 195], microscopy [3, 89], and astronomical imaging [196, 197]. The Poisson probability mass function is given by

$$p(n) = \frac{\lambda^n e^{-\lambda}}{n!}, \quad n = 0, 1, 2, \ldots \tag{6.46}$$

with $\lambda > 0$ being a rate parameter of the distribution, which is equal to both its mean and variance. In "Poissonian" imaging it is standard to associate λ with the values of an original noise-free image X. In this case, the conditional density $p(Y_k \mid X_k)$ has the form of

$$p_{Y_k|X_k}(y \mid x) = \begin{cases} \frac{e^{-x}\, x^y}{y!}, & y = 0, 1, 2, \ldots \\ 0, & \text{otherwise}, \end{cases} \tag{6.47}$$

Proposition 6.6.3. *The CSSM C^S and CRSM C^R in the case of Poisson noise contamination can be shown to be equal to*

$$C^S_{s,t,k}(y_1, y_2) = \frac{\Gamma(y_1 + y_2 + 1)}{\sqrt{\Gamma(2y_1 + 1)\Gamma(2y_2 + 1)}} \tag{6.48}$$

and

$$C^R_{s,t,k}(y_1, y_2) = \frac{\Gamma(y_1 + y_2 + 2)}{\sqrt{\Gamma(2y_1 + 2)\Gamma(2y_2 + 2)}}, \tag{6.49}$$

respectively.

The derivations are provided in Appendix B.3. We also note that, for the case of Poisson noises, the WML-optimal formulation of NLM averaging is given by (6.1).

6.6.4 Rician Distribution

The Rician noise contamination is important in a number of imaging application, a particularly important instance of which is Magnetic Resonance Imaging (MRI) [198, 199, 66]. The Rician probability density is given by

$$p_M(m) = \begin{cases} \frac{m}{\sigma^2} \exp\left(\frac{-(m^2+A^2)}{2\sigma^2}\right) I_0\left(\frac{m\,A}{\sigma^2}\right), & m \geq 0 \\ 0, & m < 0 \end{cases} \tag{6.50}$$

where $A \geq 0$ and σ^2 are distribution parameters (with σ^2 accounting for the level of noise) and I_0 is the zero order modified Bessel function of the first kind. Moreover, in the case of

Rician noises, it is standard to associate A with the values of the original image X, which results in the conditional density $p_{Y_k|X_k}(y \mid x)$ defined as

$$p_{Y_k|X_k}(y \mid x) = \begin{cases} \frac{y}{\sigma^2} \exp\left(\frac{-(y^2+x^2)}{2\sigma^2}\right) I_0\left(\frac{y\,x}{\sigma^2}\right), & y \geq 0 \\ 0, & y < 0 \end{cases} \tag{6.51}$$

Unfortunately, for the case of (6.51), we have not been able to find a closed-form expression for the CSSM C^S, which leaves numerical integration as an alternative practical option for its computation. Such a workaround, however, would adversely affect the computational efficiency of the resulting NLM filter, and this is why we advice against choosing this option. On the other hand, a closed-form expression for the CRSM C^R is readily available.

Proposition 6.6.4. *In the case of Rician noise contamination, the CRSM can be shown to be given by*

$$C^R_{s,t,k}(m_1, m_2) = \frac{I_0\left(\frac{m_1 m_2}{2\sigma^2}\right)}{\sqrt{I_0\left(\frac{m_1^2}{2\sigma^2}\right) I_0\left(\frac{m_2^2}{2\sigma^2}\right)}}. \tag{6.52}$$

The derivation is provided in Appendix B.4.

One property of the CRSM above is particularly worth mentioning. Specifically, in the case of high SNR, i.e. when $m_1, m_2 \gg \sigma^2$, the Bessel function I_0 can be approximated as

$$I_0(z) \approx \frac{e^z}{\sqrt{2\pi z}}. \tag{6.53}$$

Subsequently, plugging this approximation into (6.52) results in

$$C^R_{s,t,k}(m_1, m_2) \approx \exp\left(-\frac{(m_1 - m_2)^2}{4\sigma^2}\right), \tag{6.54}$$

which is identical to the Gaussian similarity measure as given by (6.5). Such an observation is important because it is known that the Rician density tends towards Gaussian for high SNR [198]. Thus, it appears that the above limiting property is "inherited" by the SM (6.52) as well.

Unfortunately, in the case of Rician noise contamination, straightforward application of the WML approach (6.27) does not yield a closed-form expression for NLM averaging; thereby necessitating the use of iterative maximization. It goes without saying that such an

iterative scheme would substantially increase the computational burden of NLM filtering, and it therefore should be avoided in practical settings. A particularly elegant method for overcoming the above difficulty was proposed in [66, 200, 173] based on the fact that the expected value of the square of a Rician random variable M in (6.50) bears a simple relation to its parameters A and σ^2, viz.

$$E(M^2) = A^2 + 2\sigma^2. \tag{6.55}$$

Consequently, this fact makes it possible to compute the estimate $\hat{X}_s$ according to

$$\hat{X}_s = \sqrt{\max\left(\left(\frac{\sum_t w_{s,t} Y_t}{\sum_t w_{s,t}}\right)^2 - 2\sigma^2, 0\right)}, \tag{6.56}$$

where the max operation is used to guarantee $\hat{X}_s$ to be real and positive valued. In regard to the above expression it is worthwhile pointing out that, even though $\left(\sum_t w_{s,t} Y_t / \sum_t w_{s,t}\right)^2$ does not look like a *bona fide* estimate of $E(Y_s^2)$, it has been demonstrated to produce accurate and stable reconstruction results (see, e.g., [66]) – this fact will be further supported in the experimental part of this chapter.

6.6.5 Non-Central Chi Square Distribution with Two Degrees of Freedom

As it was mentioned above, the mean value of a Rician random variable does not bear a simple relationship with the parameter A in (6.50) which represents the quantity of interest to be recovered. On the other hand, the mean of the square of the random variable does have a simple relation with A^2 as given by (6.55). This fact has motivated researches to develop denoising schemes which process the squared values of Rician images instead of the original data [201, 173].

Let M be a Rician random variable with probability density given by (6.50), and let G be the square of its σ-normalized version defined as $G \triangleq M^2/\sigma^2$. Then, G can be shown to follow a non-central chi square distribution (NCCS) with two degrees of freedom, whose pdf is given by

$$p_G(g) = \begin{cases} \frac{1}{2} e^{-\frac{g+F}{2}} I_0(\sqrt{g\,F}), & g \geq 0 \\ 0, & g < 0, \end{cases} \tag{6.57}$$

where $F \triangleq A^2/\sigma^2$, with A being the same parameter as in (6.50). Consequently, subject to the renormalization $\tilde{X}_k = X_k^2/\sigma^2$ and $\tilde{Y}_k = Y_k^2/\sigma^2$, the conditional density $p_{\tilde{Y}_k|\tilde{X}_k}(\tilde{y}|\tilde{x})$ has the form of

$$p_{\tilde{Y}_k|\tilde{X}_k}(\tilde{y} \mid \tilde{x}) = \begin{cases} \frac{1}{2} e^{-\frac{\tilde{y}+\tilde{x}}{2}} I_0(\sqrt{\tilde{y}\,\tilde{x}}), & \tilde{y} \geq 0 \\ 0, & \tilde{y} < 0. \end{cases} \tag{6.58}$$

Unfortunately, for the case of (6.58), the CRSM C^R does not admit a closed-form solution, while its computation via numerical integration would make the resulting NLM filter of little practical use due to its low computational efficiency. On the other hand, the CSSM C^S does have a closed-form expression which is given in the following proposition.

Proposition 6.6.5. *The CSSM in the case of NCCS noise contamination is given by*

$$C^S_{s,t,k}(\tilde{y}_1, \tilde{y}_2) = \frac{I_0\left(\sqrt{\tilde{y}_1 \tilde{y}_2}/2\right)}{\sqrt{I_0\left(\tilde{y}_1/2\right) I_0\left(\tilde{y}_2/2\right)}}. \tag{6.59}$$

The derivation is provided in Appendix B.5.

Once again, one can observe that, for sufficiently high values of SNR, replacing the Bessel function I_0 by its approximation from (6.53) will result in "gaussianization" of the CSSM C^S, which in this case simplifies to

$$C^S_{s,t,k}(\tilde{y}_1, \tilde{y}_2) \approx \exp\left(-\frac{(\sqrt{\tilde{y}_1} - \sqrt{\tilde{y}_2})^2}{4}\right). \tag{6.60}$$

This observation is particularly important in view of the fact that $\sigma\sqrt{G}$ follows a Rician density, which converges to a Gaussian pdf as SNR increases. Moreover, it is interesting to point out that the expressions for C^S for NCCS distribution and C^R for Rician distribution are equivalent under the substitution $G = (M/\sigma)^2$, which is precisely the relation between G and M. In other words, in terms of the values of C^S and C^R, two patches of G image and its corresponding M image are equally similar.

Finally, in the case of NCCS statistics, a practical method of NLM averaging can be derived using the relation in (6.55), in which case an estimate $\hat{X}_s$ of X_s can be computed according to

$$\hat{X}_s = \sigma\sqrt{\max\left(\frac{\sum_t w_{s,t} Y_t^2}{\sum_t w_{s,t}} - 2, 0\right)}. \tag{6.61}$$

where Y is the image contaminated by Rician noise and X is the underlying noise-free image. For the convenience of referencing, Table 6.1 provides a summary of the proposed SMs associated with different statistics of noises.

Table 6.1: The proposed similarity measures for different statistics of noises

	SM	
Noise statistics	C^S	C^R
Additive Gaussian	$\exp\left(-\frac{\lvert y_1-y_2\rvert^2}{4h^2}\right)$	–
Multiplicative Rayleigh	$\left(\frac{y_1y_2}{y_1^2+y_2^2}\right)^{3/2}$	$\frac{y_1y_2}{y_1^2+y_2^2}$
Multiplicative Gamma	$\frac{(y_1y_2)^{m-\frac{1}{2}}}{(y_1+y_2)^{2m-1}}$	$\frac{(y_1y_2)^{m-1}}{(y_1+y_2)^{2m-2}}$
Poisson	$\frac{\Gamma(y_1+y_2+1)}{\sqrt{\Gamma(2y_1+1)\Gamma(2y_2+1)}}$	$\frac{\Gamma(y_1+y_2+2)}{\sqrt{\Gamma(2y_1+2)\Gamma(2y_2+2)}}$
NCCS	$\frac{I_0(0.5\sqrt{y_1y_2})}{\sqrt{I_0(0.5y_1)I_0(0.5y_2)}}$	–
Rician	–	$\frac{I_0\left(\frac{y_1y_2}{2\sigma^2}\right)}{\sqrt{I_0\left(\frac{y_1^2}{2\sigma^2}\right)I_0\left(\frac{y_2^2}{2\sigma^2}\right)}}$

6.7 Results

The main contribution of this chapter has been to present the formal theory of FCS as well as a unified theoretical framework for the derivation of various SMs in order to obtain the fuzzy clustering for different types of noise. Also, wherever relevant, the present work has connected the obtained results to various SMs which have been already proposed in the literature. In this section, some principal experimental results have been demonstrated in order to evaluate the efficiency of the proposed similarity measures. The results will be compared with those produced by a number of state-of-the-art filtering techniques, which do not necessarily fall in the category of NLM filters. We will, however, skip addressing the case of Gaussian noise contamination (which has been studied innumerable times) and concentrate instead on less conventional noise types.

In what follows, the performance of the algorithms will be assessed in terms of *signal to noise ratio* (SNR) defined as

$$\text{SNR (in dB)} = 20\log_{10}\frac{\|X\|_{\mathbb{F}}}{\|X-\hat{X}\|_{\mathbb{F}}}, \tag{6.62}$$

where X and $\hat{X}$ represent the original image and its estimate, respectively and $\|\cdot\|_{\mathbb{F}}$ represents the Frobenius matrix norm. As a different performance metric, the SSIM index [202] will be reported as well.

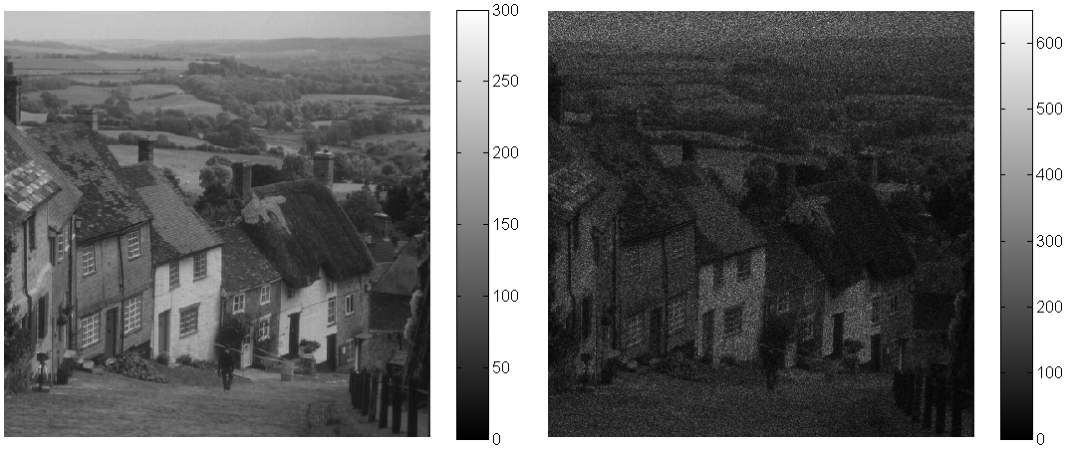
6.7.1 Multiplicative Rayleigh Noise

To demonstrate the result of denoising of images contaminated with multiplicative Rayleigh noise, we consider the standard "cameraman" image shown in Fig. 6.1(a). For this purpose, the image was contaminated by multiplicative Rayleigh noise ($\sigma = 1$), with a typical noisy image depicted in Fig. 6.1(b). Since for the case at hand the CSSM C^S and CRSM C^R are equivalent (i.e., producing the same weights through appropriately adjusting the NLM parameter α), it is sufficient to demonstrate the results for C^S only. To this end, the NLM filter has been implemented with a patch size of 5×5, a search window size of 21×21 and $\alpha = 2.5$, whereas $\beta_k,\ k \in I$ have been chosen to correspond to a Gaussian mask with standard deviation equal to 2. The result of image denoising using NLM filtering with the proposed C^S is shown in Fig. 6.1(d). The reconstruction has been compared with the method proposed in [30], which performs Bayesian estimation of the original image using total variation regularization. The regularization parameter of the reference algorithm has been chosen so as to maximize the SNR of its resulting image, with an example of the resulting reconstruction shown in Fig. 6.1(c). The SNR and the SSIM indices produced by both methods under comparison are indicated in the caption of Fig. 6.1, from which one can see that the NLM filtering yields a better reconstruction in terms of both quality measures.

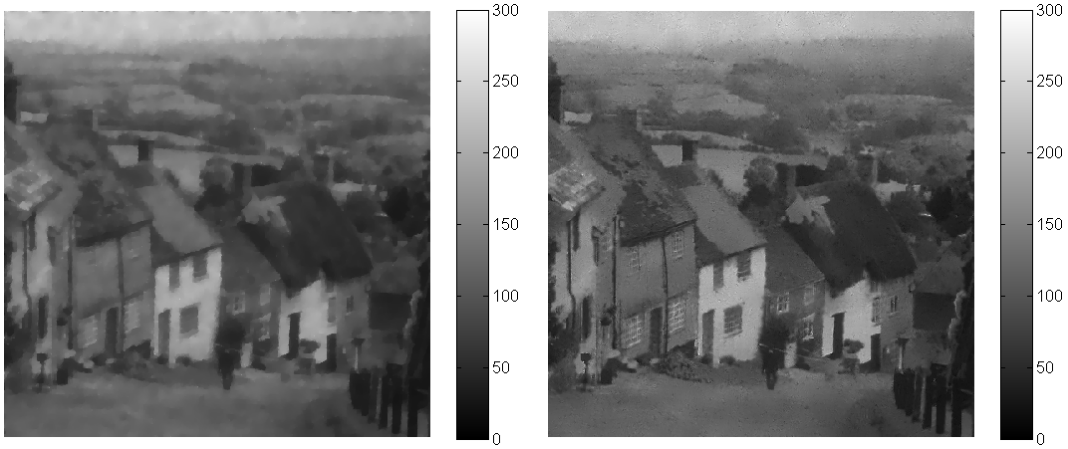
6.7.2 Multiplicative Gamma Noise

As the next validation step, the case of multiplicative Gamma noise contamination is considered in this section. In this case, the "hill" image shown in Fig. 6.2(a) was used as the original scene to be recovered. For the purpose of the experiment, the image was contaminated with multiplicative Gamma noise ($m = 12$), with a typical example of the resulting noisy image shown in Fig. 6.2(b). Once again, since for the case at hand the CSSM C^S and CRSM C^R are equivalent, only the results of the C^S-based NLM filtering are reported in this section. In particular, the NLM filter was implemented with a 7×7 patch size, a 29×29 search window, and $\alpha = 1/3$. The tuning coefficients $\beta_k,\ k \in I$ have been chosen to correspond to a Gaussian mask of standard deviation 2. Also, the method of [30] was used as a reference approach. The regularization parameter of the reference method was set to approach the maximum SNR of its resulting estimates. The results of the compared methods are shown in Fig. 6.2, with their respective SNR and SSIM indices indicated in the caption of the figure. One can see that the proposed NLM filter outperforms the reference method in terms of both SNR and SSIM values.

Figure 6.1: (a) Original image; (b) Noisy image (SNR = 3.1018 dB, SSIM = 0.2191); (c) Reconstruction using the method of [30] (SNR = 17.1836 dB, SSIM = 0.7134), and (d) Reconstruction using NLM filtering with the proposed C^S (SNR = 18.1878 dB, SSIM = 0.7176).



(a) (b)



(c) (d)

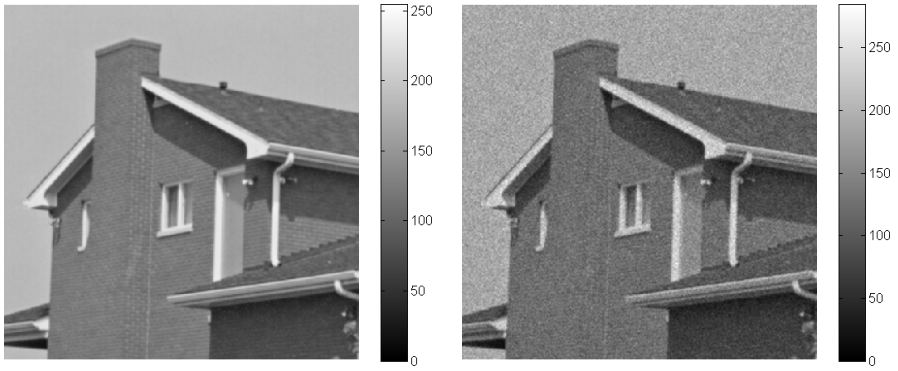
Figure 6.2: (a) Original image; (b) Noisy image (SNR = 10.7971 dB, SSIM = 0.2920); (c) Reconstruction obtained using the reference method of [30] (SNR = 20.5838 dB, SSIM = 0.6636); (d) Reconstruction using NLM filtering with the proposed C^S (SNR = 21.3100 dB, SSIM = 0.6953).

6.7.3 Poissonian Noise

To investigate the performance of image denoising in the case of "Poissonian" noises, a different test image has been employed. In particular, the original image has been selected to be the standard "house" image. The test image along with its noisy counterpart are shown in Fig. 6.3(a) and Fig. 6.3(b), respectively, while the reconstructions obtained using the C^S and C^R-based NLM filtering are shown in subplots Fig. 6.3(d) and Fig. 6.3(e), respectively. In both cases, the NLM filter was implemented with a patch size of 5×5, a search window size of 11×11, and $\alpha = 0.5$. Also, in both cases, the tuning coefficients $\beta_k,\ k \in I$ were chosen to correspond to a Gaussian mask of standard deviation 2. At the same time, the reconstruction result obtained using the reference method of [22] is depicted in Fig. 6.3(c). The SNR and SSIM values obtained with the different methods under comparison are indicated in the captions of the same figure, and they indicate that the NL filters outperform the Bayesian approach of [22] in terms of these two measures.

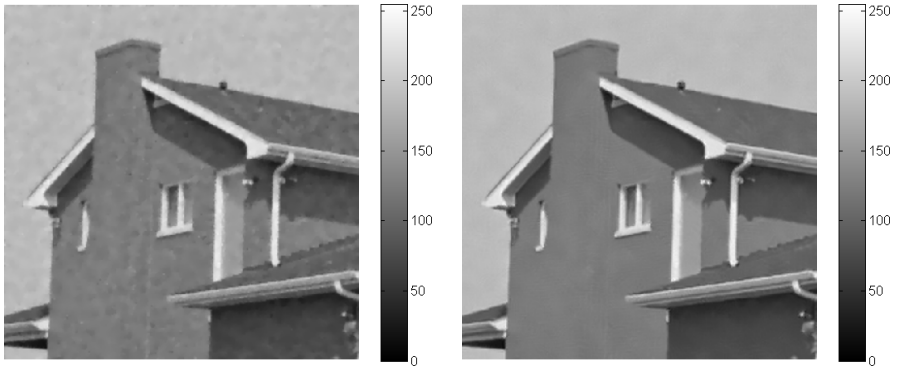
6.7.4 Rician and NCCS Noises

In this subsection, we demonstrate the results of NLM filtering of medical MRI images contaminated by Rician and NCCS noises. The test image used for the purpose of this numerical experiment is shown in Fig. 6.4(a) and represents a standard T1 image of an axial cranial cross-section. The noisy version of the image shown in 6.4(b) has been obtained by contaminating the original image with Rician noise ($\sigma = 12$). Both the C^S-and C^R-based NLM filters have been implemented using a 5×5 neighbourhood, an 11×11 search window, while α was set to be equal to 0.5. The turning weights β_k in (6.4) were defined to correspond to a Gaussian window of standard deviation 0.5. The performance of the proposed algorithms have also been compared with that of two reference methods, viz. the wavelet denoising method (WDN) of [49] and the unbiased NLM (UNLM) filtering method of [66]. Both reference methods have been implemented using codes available at the authors' websites at `http://telin.ugent.be/~sanja/Sanja_files/Software/` and `http://personales.upv.es/jmanjon/denoising/nlm2d.htm`, respectively. The reconstruction results obtained using all the methods under comparison are shown in Fig. 6.4, with their associated SNR and SSIM values indicated in the caption of the figure. Once again, one can see that the proposed denoising algorithms outperform the reference methods in terms of both the performance criteria.



(a)

(b)



(c)

(d)

(e)

Figure 6.3: (a) Original image; (b) Noisy image (SNR = 21.8949 dB, SSIM = 0.5550); (c) Reconstruction obtained using the method of [22] (SNR = 28.5426 dB, SSIM = 0.8709); (d) Reconstruction obtained using the C^S-based NLM filtering (SNR = 29.9788 dB, SSIM = 0.8830); (e) Reconstruction obtained using the C^R-based NLM filtering (SNR = 29.9788 dB, SSIM = 0.8830).

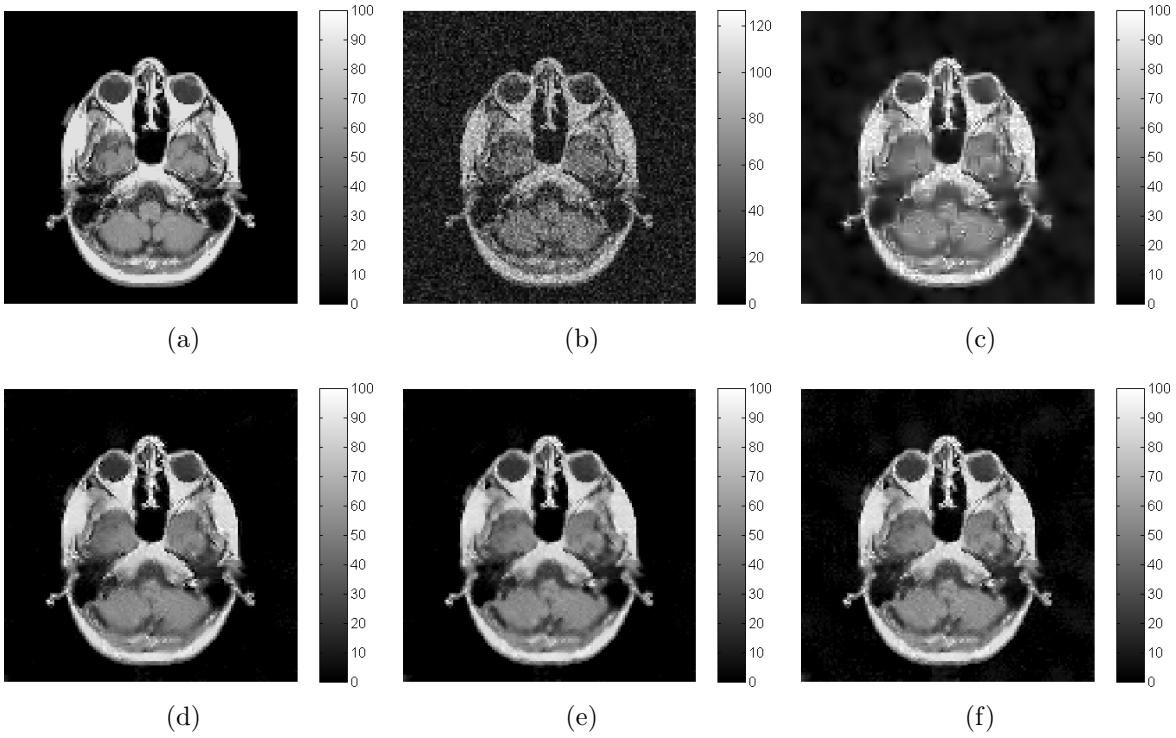


Figure 6.4: (a) Original image; (b) Noisy image (SNR: 13.5556 dB, SSIM: 0.8771); (c) Reconstruction obtained using WDN (SNR: 15.0848 dB, SSIM: 0.8971); (d) Reconstruction obtained using UNLM (SNR: 15.7137 dB, SSIM: 0.9008); (e) Reconstruction obtained using the proposed C^R-based NLM filter (SNR: 16.8630 dB, SSIM: 0.9291), and (f) Reconstruction obtained using the proposed C^S-based NLM filter (SNR: 16.9115 dB, SSIM: 0.9283).

Chapter 7

Summary and Future Directions

The thesis has extended the concept and applicability of variable splitting as an efficient solution to the problem of image reconstruction. Specifically, it has addressed the fact that several image reconstruction applications of modern days are associated with acquisition and processing of large amount of data that increases the computational burden of the reconstruction algorithms proportionally. Consequently, in many practical applications, the key challenge is to design a computationally efficient algorithm while guaranteeing acceptable reconstruction accuracy. The thesis has provided a new insight in this direction for addressing many real life problems. In particular, we have considered two important methods under the class of variable splitting algorithms. Subsequently, the methods have been used in solving a number of problems which are very important in modern day image processing.

7.1 Contributions

The summary of the contributions of the thesis, while highlighting its importance, is presented below.

7.1.1 Compressed Sensing using Cross-Domain Constraints

The thesis has introduced a new class of problems for multidimensional vector valued images that require acquisition of a large amount of data which is often not possible due to practical constraints. Such a scenario is prevalent in a number of important practical

applications. As a solution to the problem, we have proposed to acquire reduced amount of data and subsequently reconstruct the quantity of interest using the theory of compressed sensing. In order to improve the accuracy of the reconstruction, we have formulated an optimization problem which incorporates cross-domain composite constraints. Because of the large dimensionality of the data as well as the nature of the composite constraints, direct solution of the optimization problem requires significant computational resources and hence is not desired. As a solution to the latter problem, we have proposed to use the tools of variable splitting to simplify the reconstruction steps. In particular, the solution involves iterative application of some simpler optimization problems whose fast solutions are available in the literature. As an application of the proposed framework, we have considered the problem of reconstruction of HARDI signals from subcritical samples. Specifically, the acquisition of HARDI data at each voxel requires a large number of samples which translates to a relatively long acquisition time that impairs the practical value of this imaging modality. As a practical solution, it has been suggested to acquire fewer samples while reconstructing the data using the proposed method. In particular, we have been able to obtain reasonable reconstruction using only 16-20 diffusion encoding directions as opposed to 60-100 directions typically used in HARDI acquisition.

7.1.2 Hybrid Blind Deconvolution

In many practical applications, the acquired image is blurred due to the attenuation of high frequency components. The process of image deconvolution aims at recovering the attenuated frequency components in order to result in a sharper image with enhanced details. The problem is more challenging when the deconvolution process is blind, i.e., when the point spread function is not known in advance. A specific type of blind deconvolution algorithm is hybrid blind deconvolution. The thesis has provided an efficient numerical solution to that approach using the concept of variable splitting. The hybrid deconvolution technique aims at finding an optimal inverse filter using a two stage reconstruction procedure. In the first stage, some partial information about the point spread function of the imaging system (namely, its magnitude spectrum) is recovered. Subsequently, the obtained information is exploited to explicitly constrain the procedure of inverse filtering. The latter is realized in the form of an optimization problem which couples regularizers defined over the spatial as well as the frequency domain. Direct solution of the optimization problem is non-trivial because of the coupling of the regularization terms defined over different domains of definition of the image. On the other hand, the proposed method leads to a particularly efficient numerical scheme, which can be implemented as a succession of analytically computable proximity operations. As an example, we have shown the appli-

cation of the method to the problem of blind deconvolution of medical ultrasound images, which are, in general, characterized by low resolution. In particular, we have shown how the proposed method can result in improvement of resolution and contrast of the images.

7.1.3 Fuzzy Clustering Splitting

The thesis has addressed the problem of image denoising which is a very important problem in different fields of image processing and computer vision. In this aspect, the contribution of the thesis is two-fold. First, the thesis has formally proposed the principles of fuzzy clustering splitting for denoising of images. In particular, the proposed approach splits a denoising problem into two sub-problems; fuzzy clustering of the pixels based on their similarities and estimation of the underlying intensities based on the fuzzy clustering. Subsequently, it has also been shown how the fuzzy clustering splitting approach generalizes the concept of neighbourhood filtering. As a second contribution, the thesis has proposed a unifying framework of defining the similarity measure between two pixel-intensities within an image which is required to perform the fuzzy clustering. It has also discussed the theoretical advantages of the proposed measure over existing measures. Thereafter, it has provided closed form expression of the similarity measure in the case of several types of measurement noise that are prevalent in a number of important imaging modalities. Finally, the quality of reconstruction using the proposed similarity measures has been demonstrated through a number of numerical experiments along with comparisons with existing alternative techniques.

7.2 Future Directions

The thesis can lead to several future researches, some of which are mentioned below.

1. The approach of compressed sensing presented in this thesis can be applied to problems involving multidimensional vector valued images. In particular, a number of important applications, namely hyperspectral imaging and sensor networks, have been mentioned in this thesis where the proposed approach can be applied. The challenges involve designing appropriate acquisition and reconstruction bases depending on the nature of the data to be acquired. Subsequently, the theory of compressed sensing can be applied while the computational complexity can be significantly reduced using the proposed approach.

2. It provides the scope of extending the concept of hybrid blind deconvolution to the case of general imagery. In fact, in the case of other types of images, one needs to design efficient methods for obtaining the magnitude spectrum of point spread function. In this context, however, the thesis presents a wider scope for future research. In particular, it can be used to solve problems where one has access to partial information of the unknown quantity of interest. Subsequently, the complete information can be obtained by constraining the optimization problem using the partial information. The constraints can be tackled separately using the proposed approach, thus leading to a significant reduction in computational complexity.

3. The denoising approach proposed in this thesis can remove independent and identically distributed random noise which has well defined statistical characteristics. This assumption is often not valid in different imaging modalities. For example, in the case of MRI data, there can be various structural artifacts which cause further degradation of the image. Such degradations cannot be addressed using the proposed fuzzy clustering splitting . Accordingly, the technique needs to be extended in the future to incorporate the aspect of denoising structural noise.

APPENDICES

Appendix A

Proof of Proposition 5.4.1

In order to solve the complex-valued problem,

$$\arg\min_{z} \frac{1}{2}(|z|a-1)^2 + \frac{\mu}{2}|zB-C|^2 \tag{A.1}$$

where $a, \mu \in \mathbb{R}_+$, and $B, C \in \mathbb{C}$ let us first find a minimizer of

$$E(x) = \frac{1}{2}(|x|\,a-1)^2 + \frac{\mu}{2}(x\,b-c)^2 \tag{A.2}$$

where $x, b, c \in \mathbb{R}$, while $a, \mu \in \mathbb{R}_+$. The function in (A.2) is not convex and hence one can obtain an extremum by equating the derivative to zero. Accordingly, for $x \neq 0$, the condition for extrema is given by

$$(|x|a-1)a\,\mathrm{sign}(x) + \mu b(xb-c) = 0 \tag{A.3}$$

which is equivalent to

$$x - \frac{a}{a^2+\mu b^2}\mathrm{sign}(x) = \frac{\mu bc}{a^2+\mu b^2} \tag{A.4}$$

On the other hand, for $x = 0$, the function in (A.2) is not differentiable and hence $x = 0$ is an extremum if

$$0 \in (|x|\,a-1)\,a \cdot [-1,1] + \mu\,b(x\,b-c) \tag{A.5}$$

and using $x = 0$, this amounts to

$$\frac{\mu bc}{a^2+\mu b^2} \in \frac{a}{a^2+\mu b^2} \cdot [-1,1] \tag{A.6}$$

Hence, the solution(s) of (A.4) is an extremum (are the extrema) of $E(x)$ in (A.2). In addition, $x = 0$ is also an extremum if the condition in (A.6) is satisfied. In order to solve for x, we follow a procedure similar to the derivation of the soft thresholding function in subsection 3.2.2. In particular, denoting $\gamma = \frac{a}{a^2+\mu b^2}$, we define a multivalued function $F(x)$ (shown in Fig. A.1(a)) as given by

$$F(x) = \begin{cases} x - \gamma \text{sign}(x), & x \neq 0 \\ [-\gamma, \gamma], & x = 0 \end{cases} \tag{A.7}$$

Subsequently, the solution(s) can be obtained by finding the intercept where $F(x) = \frac{\mu bc}{a^2+\mu b^2}$. Alternatively, one can construct the inverse function of $F(x)$ and evaluate the function at the point $\frac{\mu bc}{a^2+\mu b^2}$. In particular, denoting the inverse function by I, the solution x can be obtained by

$$x = I\left(\frac{\mu bc}{a^2 + \mu b^2}\right) \tag{A.8}$$

To obtain the inverse function I, we change the coordinates of the plot of $F(x)$ in Fig. A.1(a) and the resulting plot is shown in Fig. A.1(b). Formally, the inverse function I can be defined as

$$I(\tau) = \begin{cases} \tau - \gamma, & \tau < -\gamma \\ \{0, -2\gamma\}, & \tau = -\gamma \\ \{\tau - \gamma, 0, \tau + \gamma\}, & -\gamma < \tau < \gamma \\ \{0, 2\gamma\}, & \tau = \gamma \\ \tau + \gamma, & \tau > \gamma \end{cases} \tag{A.9}$$

The fact that I is multivalued over the interval $[-\gamma, \gamma]$ provides multiple solutions for x, each of which is a local extremum. In order to obtain the minimizer, we resubstitute the solutions given by (A.8) into the original cost function in (A.2) to find which of the candidate solutions results in a smaller value of the cost function. In particular, the three solutions corresponding to the case $\frac{\mu bc}{a^2+\mu b^2} \in [-\gamma, \gamma]$ are given by

$$x = \begin{cases} 0 \\ \frac{\mu bc}{a^2+\mu b^2} + \gamma \\ \frac{\mu bc}{a^2+\mu b^2} - \gamma \end{cases}. \tag{A.10}$$

Substituting $\gamma = \frac{a}{a^2+\mu b^2}$, the values of the cost function corresponding to the values of x in

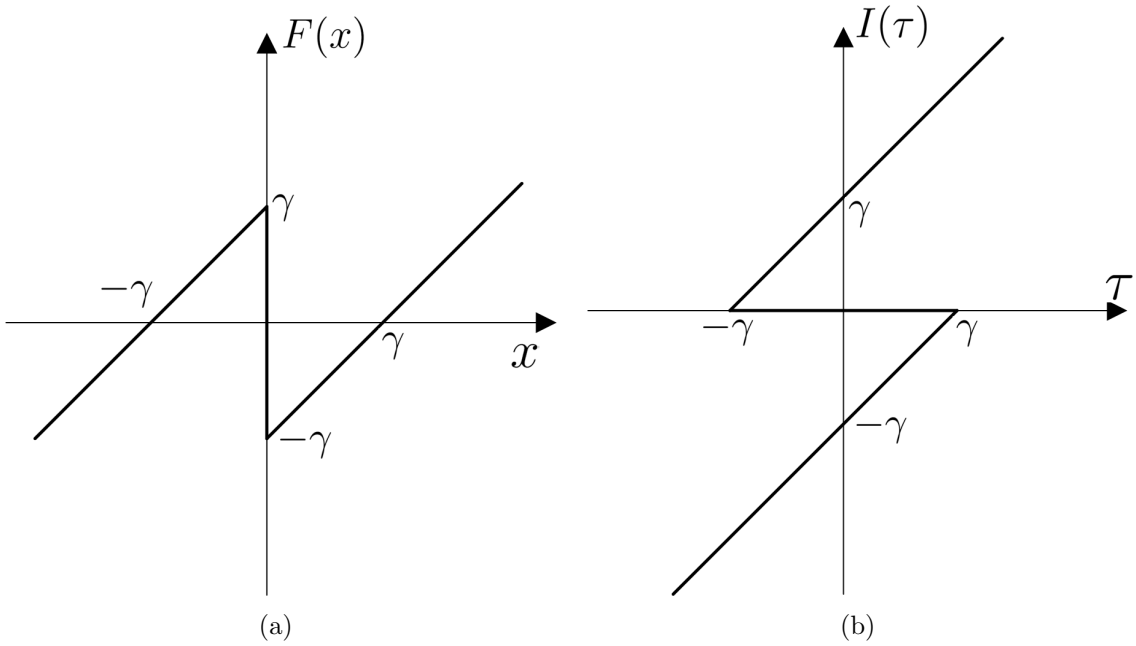


Figure A.1: (a) Plot of $F(x)$ in (A.7) and (b) its inverse function in (A.9)

(A.10) are given by

$$E(x) = \begin{cases} \frac{1}{2} + \frac{\mu c^2}{2}, & x = 0 \\ \frac{1}{2}\left(|\mu bc + a|\,\frac{a}{a^2+\mu b^2} - 1\right)^2 + \frac{\mu}{2}\left((\mu bc + a)\,\frac{b}{a^2+\mu b^2} - c\right)^2, & x = \frac{\mu bc+a}{a^2+\mu b^2} \\ \frac{1}{2}\left(|\mu bc - a|\,\frac{a}{a^2+\mu b^2} - 1\right)^2 + \frac{\mu}{2}\left((\mu bc - a)\,\frac{b}{a^2+\mu b^2} - c\right)^2, & x = \frac{\mu bc-a}{a^2+\mu b^2} \end{cases} \tag{A.11}$$

To compare the values of $E(x)$ for the different values of x, we use $\frac{\mu bc}{a^2+\mu b^2} \in [-\gamma, \gamma]$ and $\gamma = \frac{a}{a^2+\mu b^2}$ to obtain

$$-2a \le \mu bc - a \le 0, \text{ and } 0 \le \mu bc + a \le 2a \tag{A.12}$$

which suggest that $\mu bc - a$ is negative and $\mu bc + a$ is positive. Subsequently, this leads to

$$|\mu bc + a| = \mu bc + a, \text{ and } |\mu bc - a| = -(\mu bc - a) \tag{A.13}$$

Using these facts, it can be shown that

$$\frac{1}{2} + \frac{\mu c^2}{2} > \frac{1}{2}\left(|\mu bc + a|\,\frac{a}{a^2+\mu b^2} - 1\right)^2 + \frac{\mu}{2}\left((\mu bc + a)\,\frac{b}{a^2+\mu b^2} - c\right)^2 \tag{A.14}$$

and

$$\frac{1}{2}+\frac{\mu c^2}{2} > \frac{1}{2}\left(|\mu bc-a|\,\frac{a}{a^2+\mu b^2}-1\right)^2+\frac{\mu}{2}\left((\mu bc-a)\,\frac{b}{a^2+\mu b^2}-c\right)^2 \tag{A.15}$$

Moreover, it can also be shown that

$$E\left(\frac{\mu bc+a}{a^2+\mu b^2}\right) \gtreqless E\left(\frac{\mu bc-a}{a^2+\mu b^2}\right), \quad \text{iff } bc \lesseqgtr 0 \tag{A.16}$$

Combining all the facts, the minimizer of $E(x)$ in (A.2) is given by

$$x=\begin{cases}\frac{\mu bc-a}{a^2+\mu b^2}, & bc<0\\ \frac{\mu bc+a}{a^2+\mu b^2}, & bc\geq 0\end{cases} \tag{A.17}$$

or, using the rampstep function of (5.23), the solution is given by

$$x=\text{rampstep}_{\frac{a}{a^2+\mu b^2}}\frac{\mu bc}{a^2+\mu b^2} \tag{A.18}$$

Next we consider minimizing the functional

$$E(z)=\frac{1}{2}(|z|a-1)^2+\frac{\mu}{2}|zB-C|^2$$

where $a\in\mathbb{R}_+$, and $z,B,C\in\mathbb{C}$. Let $z=x+\imath y\equiv[x,y]$, $B=B_R+\imath B_I\equiv[B_R,B_I]$, and $C=C_R+\imath C_I\equiv[C_R,C_I]$. Then

$$\begin{aligned}E(x,y)&=\frac{1}{2}(\sqrt{x^2+y^2}\,a-1)^2+\frac{\mu}{2}|(x+\imath y)(B_R+\imath B_I)-(C_R+\imath C_I)|^2\\&=\frac{1}{2}(\sqrt{x^2+y^2}\,a-1)^2+\frac{\mu}{2}(xB_R-yB_I-C_R)^2+\frac{\mu}{2}(yB_R+xB_I-C_I)^2\end{aligned}$$

Therefore

$$\begin{aligned}\frac{\partial E}{\partial x}&=(\sqrt{x^2+y^2}\,a-1)\,\frac{ax}{\sqrt{x^2+y^2}}+\mu B_R(xB_R-yB_I-C_R)+\mu B_I(yB_R+xB_I-C_I)\\&=a^2x-\frac{ax}{\sqrt{x^2+y^2}}+\mu xB_R^2-\mu yB_RB_I-\mu B_RC_R+\mu yB_IB_R+\mu xB_I^2-\mu B_IC_I\\&=a^2x-\frac{ax}{\sqrt{x^2+y^2}}+\mu x(B_R^2+B_I^2)-\mu(B_RC_R+B_IC_I)\\&=(a^2+\mu|B|^2)x-\frac{ax}{|z|}-\mu.\Re\{C.\bar{B}\}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial E}{\partial y} &= (\sqrt{x^2+y^2}\,a - 1)\,\frac{ay}{\sqrt{x^2+y^2}} - \mu B_I(xB_R - yB_I - C_R) + \mu B_R(yB_R + xB_I - C_I) \\ &= a^2 y - \frac{ay}{\sqrt{x^2+y^2}} - \mu x B_I B_R + \mu y B_I^2 + \mu B_I C_R + \mu y B_R^2 + \mu x B_R B_I - \mu B_R C_I \\ &= a^2 y - \frac{ay}{\sqrt{x^2+y^2}} + \mu y(B_R^2 + B_I^2) - \mu(B_R C_I - B_I C_R) \\ &= (a^2 + \mu|B|^2)y - \frac{ay}{|z|} - \mu.\Im\{C.\bar{B}\}\end{aligned}$$

Therefore

$$\begin{aligned}\nabla E &= (a^2 + \mu|B|^2)z - a\frac{z}{|z|} - \mu(C.\bar{B}) = 0 \\ \implies z - \frac{a}{a^2 + \mu|B|^2}\frac{z}{|z|} &= \frac{\mu C.\bar{B}}{a^2 + \mu|B|^2}\end{aligned}$$

Since $a/(a^2 + \mu|B|^2)$ is real, and z and $z/|z|$ have the same phase, the phase of the right hand side is equal to the phase of z up to a negative sign (phase offset of π). Hence, denoting

$$\alpha = \frac{a}{a^2 + \mu|B|^2}, \quad \beta = \frac{\mu C\bar{B}}{a^2 + \mu|B|^2}, \quad \theta = \tan^{-1}\left(\Im\{\beta\}/\Re\{\beta\}\right), \tag{A.19}$$

we can write $z = z_R e^{j\theta}$ and $\beta = \beta_R e^{j\theta}$ where $z_R, \beta_R \in \mathbb{R}$ (can be negative). Using such a representation, the above equation becomes

$$z_R e^{j\theta} - \alpha\frac{z_R e^{j\theta}}{|z_R|} = \beta_R e^{j\theta} \tag{A.20}$$

Multiplying by $e^{-j\theta}$, we obtain

$$z_R - \alpha\text{sign}(z_R) = \beta_R \tag{A.21}$$

The solution z_R is obtained from the real case and is given by

$$z_R = \text{rampstep}_\alpha(\beta_R) \tag{A.22}$$

and hence the complete solution is given by

$$z = \text{rampstep}_\alpha(\beta e^{-j\theta})e^{j\theta} \tag{A.23}$$

Appendix B

Derivations of Similarity Measures

In the derivation of the similarity measures, two identities will be used extensively and hence we state them here.

$$\int_0^\infty \frac{1}{x^b} \exp\left(-\frac{c}{x}\right) dx = \frac{1}{c^{b-1}}\Gamma(b-1), \quad c > 0,\, b > 1, \tag{B.1}$$

$$\Gamma(z) = \int_0^\infty e^{-x} x^{z-1}\, dx. \tag{B.2}$$

where Γ is the gamma function.

B.1 Multiplicative Rayleigh Noise

To compute C^S, we use (6.33) to obtain

$$\langle L_{y_1}, L_{y_2} \rangle_{\mathbb{L}_2} = \int_0^\infty p_{Y_{s-k}|X_{s-k}}(y_1 \mid x)\, p_{Y_{t-k}|X_{t-k}}(y_2 \mid x)\, dx \tag{B.3}$$

$$= \int_{x=0}^\infty \frac{1}{x^2}\, p_{V_{s-k}}\left(\frac{y_1}{x}\right) p_{V_{t-k}}\left(\frac{y_2}{x}\right) dx. \tag{B.4}$$

Using the Rayleigh pdf from (6.34) in (B.4), we obtain

$$\langle L_{y_1}, L_{y_2} \rangle_{\mathbb{L}_2} = \frac{y_1 y_2}{\sigma^4} \int_{x=0}^\infty \frac{1}{x^4} \exp\left(-\frac{y_1^2 + y_2^2}{2\sigma^2 x^2}\right) dx,$$

which, under the substitution $t = x^2$ and $dx = \frac{dt}{2\sqrt{t}}$, becomes

$$\langle L_{y_1}, L_{y_2} \rangle_{\mathbb{L}_2} = \frac{y_1 y_2}{\sigma^4} \int_{t=0}^{\infty} \frac{1}{t^2} \exp\left(-\frac{y_1^2 + y_2^2}{2\sigma^2 t}\right) \frac{dt}{2\sqrt{t}} = \frac{\sqrt{2}}{\sigma} \Gamma\left(\frac{3}{2}\right) \frac{y_1 y_2}{(y_1^2 + y_2^2)^{3/2}}, \tag{B.5}$$

where the last step is obtained using (B.1). Consequently, taking into account the fact that

$$\|L_y\|_{\mathbb{L}_2}^2 = \frac{\sqrt{2}}{\sigma} \Gamma\left(\frac{3}{2}\right) \frac{y^2}{(2y^2)^{3/2}}, \tag{B.6}$$

we obtain

$$C_{s,t,k}^{S}(y_1, y_2) = \left(\frac{y_1 y_2}{y_1^2 + y_2^2}\right)^{3/2} = \left(\frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{-3/2}. \tag{B.7}$$

To derive C^R, we first note that

$$\langle L_{y_1}, L_{y_2} \rangle_{\mathbb{S}} = \int_{x=0}^{\infty} x \, p_{Y_{s-k}|X_{s-k}}(y_1 \mid x) p_{Y_{t-k}|X_{t-k}}(y_2 \mid x) \, dx \tag{B.8}$$

$$= \int_{x=0}^{\infty} \frac{1}{x} \, p_{V_{s-k}}\left(\frac{y_1}{x}\right) p_{V_{t-k}}\left(\frac{y_2}{x}\right) \, dx. \tag{B.9}$$

Subsequently, substitute the Rayleigh pdf in (B.9) results in

$$\langle L_{y_1}, L_{y_2} \rangle_{\mathbb{S}} = \frac{y_1 y_2}{\sigma^4} \int_{x=0}^{\infty} \frac{1}{x^3} \exp\left(-\frac{y_1^2 + y_2^2}{2\sigma^2 x^2}\right) dx, \tag{B.10}$$

which, under the substitution $t = x^2$ and $dx = \frac{dt}{2\sqrt{t}}$ and using (B.1), becomes

$$\langle L_{y_1}, L_{y_2} \rangle_{\mathbb{S}} = \frac{1}{\sigma^2} \frac{y_1 y_2}{y_1^2 + y_2^2}, \tag{B.11}$$

with the corresponding norm given by

$$\|L_y\|_{\mathbb{S}}^2 = \frac{1}{2\sigma^2}. \tag{B.12}$$

Finally, ignoring the constant terms, we obtain

$$C_{s,t,k}^{R}(y_1, y_2) = \frac{y_1 y_2}{y_1^2 + y_2^2} = \left(\frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{-1}. \tag{B.13}$$

B.2 Multiplicative Gamma Noise

To compute C^S, we substitute the Gamma pdf ((6.41)) in place of p_V in (B.4) and subsequently use (B.1) to obtain

$$\langle L_{y_1}, L_{y_2} \rangle_{\mathbb{L}_2} = m \frac{\Gamma(2m-1)}{\Gamma(m)^2} \frac{(y_1 y_2)^{m-1}}{(y_1+y_2)^{2m-1}}. \tag{B.14}$$

The norm $\|L_y\|_{\mathbb{L}_2}$ is given by

$$\|L_y\|^2_{\mathbb{L}_2} = m \frac{\Gamma(2m-1)}{\Gamma(m)^2} \frac{(y)^{2(m-1)}}{(2y)^{2m-1}} = m \frac{\Gamma(2m-1)}{\Gamma(m)^2} \frac{1}{2^{2m-1} y}. \tag{B.15}$$

Subsequently, ignoring all constant terms, C^S is given by

$$C^S_{s,t,k}(y_1, y_2) = \frac{(y_1 y_2)^{m-\frac{1}{2}}}{(y_1+y_2)^{2m-1}} = \frac{1}{\left(2 + \frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{m-\frac{1}{2}}}. \tag{B.16}$$

To obtain C^R, we substitute the Gamma pdf (6.41) in (B.9) and use (B.1) to obtain

$$\langle L_{y_1}, L_{y_2} \rangle_{\mathbb{S}} = m^2 \frac{\Gamma(2m-2)}{\Gamma(m)^2} \frac{(y_1 y_2)^{m-1}}{(y_1+y_2)^{2m-2}}. \tag{B.17}$$

In this case, the norm $\|L_y\|^2_{\mathbb{S}}$ can be found to be

$$\|L_y\|^2_{\mathbb{S}} = m^2 \frac{\Gamma(2m-2)}{\Gamma(m)^2} \frac{1}{2^{2m-1}}. \tag{B.18}$$

Therefore, after ignoring all constant terms, the expression of C^R is given by

$$C^R_{s,t,k}(y_1, y_2) = \frac{(y_1 y_2)^{m-1}}{(y_1+y_2)^{2m-2}} = \frac{1}{\left(2 + \frac{y_1}{y_2} + \frac{y_2}{y_1}\right)^{m-1}}. \tag{B.19}$$

B.3 Poisson Noise

To derive C^S for Poisson distribution, we substitute the Poisson pdf from (6.47) into (B.3) to obtain

$$\begin{aligned} \langle L_{y_1}, L_{y_2} \rangle_{\mathbb{L}_2} &= \frac{1}{y_1! y_2!} \int_0^\infty e^{-2x}\, x^{y_1+y_2}\, dx = \frac{1}{2 y_1! y_2!} \int_0^\infty e^{-\lambda} \left(\frac{\lambda}{2}\right)^{y_1+y_2} d\lambda = \\ &= \frac{1}{2^{y_1+y_2+1} y_1! y_2!}\, \Gamma(y_1+y_2+1), \end{aligned} \tag{B.20}$$

where the last line is obtained using (B.2). Consequently, $\|L_y\|^2_{\mathbb{L}_2}$ is given by

$$\|L_y\|^2_{\mathbb{L}_2} = \frac{1}{2^{2y+1}(y!)^2}\Gamma(2y+1), \tag{B.21}$$

which suggests C^S to have the form of

$$C^S_{s,t,k}(y_1, y_2) = \frac{\Gamma(y_1+y_2+1)}{\sqrt{\Gamma(2y_1+1)\Gamma(2y_2+1)}}.$$

To derive C^R for Poisson noise, we substitute the Poisson pdf in (B.8) to obtain

$$\begin{aligned}\langle L_{y_1}, L_{y_2}\rangle_{\mathbb{S}} &= \frac{1}{y_1!y_2!}\int_0^\infty x\, e^{-2x}\, x^{y_1+y_2}\, dx = \frac{1}{2^{y_1+y_2+2}y_1!y_2!}\int_0^\infty \lambda e^{-\lambda}\lambda^{y_1+y_2+1}\, d\lambda = \\ &= \frac{1}{2^{y_1+y_2+2}y_1!y_2!}\Gamma(y_1+y_2+2).\end{aligned}$$

Consequently, $\|L_y\|^2_{\mathbb{S}}$ will be given by

$$\|L_y\|^2_{\mathbb{S}} = \frac{1}{2^{2y+2}(y!)^2}\Gamma(2y+2), \tag{B.22}$$

and, therefore,

$$C^R_{s,t,k}(y_1, y_2) = \frac{\Gamma(y_1+y_2+2)}{\sqrt{\Gamma(2y_1+2)\Gamma(2y_2+2)}}.$$

B.4 Rician Distribution

To derive C^R or Rician density, we substitute the Rician pdf (6.50) into (B.8) to obtain

$$\begin{aligned}\langle L_{m_1}, L_{m_2}\rangle_{\mathbb{S}} &= \frac{m_1m_2}{\sigma^4}e^{-(m_1^2+m_2^2)/2\sigma^2}\int_0^\infty a\, e^{-a^2/\sigma^2} I_0\left(\frac{m_1a}{\sigma^2}\right) I_0\left(\frac{m_2a}{\sigma^2}\right) da = \\ &= \frac{m_1m_2}{\sigma^2}e^{-(m_1^2+m_2^2)/2\sigma^2}\int_0^\infty y\, e^{-y^2} I_0\left(\frac{m_1}{\sigma}y\right) I_0\left(\frac{m_2}{\sigma}y\right) dy = \qquad \text{(B.23)} \\ &= \frac{m_1m_2}{\sigma^2}e^{-(m_1^2+m_2^2)/2\sigma^2}\int_0^\infty y\, e^{-y^2} J_0\left(\imath\frac{m_1}{\sigma}y\right) J_0\left(\imath\frac{m_2}{\sigma}y\right) dy, \qquad \text{(B.24)}\end{aligned}$$

where (B.23) was obtained by substituting $y = a/\sigma$, while (B.24) was obtained based on the fact that $I_0(z) = J_0(\imath z)$ (where J_0 denotes the Bessel function of the first kind and zero order, and $\imath = \sqrt{-1}$). Subsequently, using formula (2.32) of [203], we have

$$\begin{aligned}\langle L_{m_1}, L_{m_2}\rangle_{\mathbb{S}} &= \frac{m_1 m_2}{2\sigma^2} e^{-[(m_1^2+m_2^2)/4\sigma^2]} I_0\left(-\frac{m_1 m_2}{2\sigma^2}\right) \\ &= \frac{m_1 m_2}{2\sigma^2} e^{-[(m_1^2+m_2^2)/4\sigma^2]} I_0\left(\frac{m_1 m_2}{2\sigma^2}\right),\end{aligned} \tag{B.25}$$

where the last equation was obtained by using $I_0(-z) = I_0(z)$. Consequently, $\|L_m\|_{\mathbb{S}}^2$ is equal to

$$\|L_m\|_{\mathbb{S}}^2 = \frac{m^2}{2\sigma^2} e^{-m^2/2\sigma^2} I_0\left(\frac{m^2}{s\sigma^2}\right), \tag{B.26}$$

and thus

$$C_{s,t,k}^R(m_1, m_2) = \frac{I_0\left(\frac{m_1 m_2}{2\sigma^2}\right)}{\sqrt{I_0\left(\frac{m_1^2}{2\sigma^2}\right) I_0\left(\frac{m_2^2}{2\sigma^2}\right)}}. \tag{B.27}$$

B.5 Non-Central Chi Square Distribution with Two Degrees of Freedom

To derive C^S for NCCS distribution, we substitute the NCCS pdf from (6.57) into (B.3) to obtain

$$\langle L_{g_1}, L_{g_2}\rangle = \frac{1}{4} e^{-(g_1+g_2)/2} \int_{x=0}^{\infty} e^{-f} I_0(\sqrt{g_1 f}) I_0(\sqrt{g_2 f})\, df. \tag{B.28}$$

Subsequently, substituting $t = \sqrt{f}$ and using $I_0(z) = J_0(\imath z)$, we obtain

$$\langle L_{g_1}, L_{g_2}\rangle = \frac{1}{2} e^{-(g_1+g_2)/2} \int_{t=0}^{\infty} t e^{-t^2} J_0(\imath\sqrt{g_1} t) J_0(\imath\sqrt{g_2} t) dt. \tag{B.29}$$

Finally, using equation (2.32) of [203], we have

$$\langle L_{g_1}, L_{g_2}\rangle = \frac{1}{4} e^{-(g_1+g_2)/4} I_0\left(-\frac{\sqrt{g_1 g_2}}{2}\right) = \frac{1}{4} e^{-(g_1+g_2)/4} I_0\left(\frac{\sqrt{g_1 g_2}}{2}\right), \tag{B.30}$$

where the fact that $I_0(-z) = I_0(z)$ was used once again. Consequently, $\|L_g\|_{\mathbb{L}_2}^2$ can be obtained as

$$\|L_g\|_{\mathbb{L}_2}^2 = \frac{1}{4} e^{-g/2} I_0\left(\frac{g}{2}\right), \tag{B.31}$$

and therefore

$$C^S_{s,t,k}(g_1, g_2) = \frac{I_0\left(\frac{\sqrt{g_1 g_2}}{2}\right)}{\sqrt{I_0\left(\frac{g_1}{2}\right) I_0\left(\frac{g_2}{2}\right)}}. \tag{B.32}$$

References

[1] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition, 2001.

[2] L. L. Bonilla, A. Carpio, O. Dorn, M. Moscoso, F. Natterer, G. C. Papanicolaou, M. L. Rapún, and A. Teta. *Inverse Problems and Imaging*. Lecture Notes in Mathematics. Springer, 2008.

[3] W. H. Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America A*, 62(1):55–59, 1972.

[4] M. A. T. Figueiredo and R. D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, 2003.

[5] 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.

[6] J.-S. Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(2):165–168, March 1980.

[7] 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), December 2009.

[8] M. Elad and A. Feuer. Superresolution restoration of an image sequence: Adaptive filtering approach. *IEEE Transactions on Image Processing*, 8(3):387–395, March 1999.

[9] 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*, volume 2, pages 60–65, 2005.

[10] 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.

[11] D. L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.

[12] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, November 2003.

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

[14] C. Oliver and S. Quegan. *Understanding synthetic aperture radar images*. SciTech Publishing, 2004.

[15] O. Michailovich and D. Adam. Phase unwrapping for 2-D blind deconvolution of ultrasound images. *IEEE Transactions on Medical Imaging*, 23(1):7–25, January 2004.

[16] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging*. CRC Press, 1998.

[17] J. B. Keller. Inverse problems. *American Mathematical Monthly*, 83:107–118, 1976.

[18] A. Tarantola. *Inverse Problem Theory Methods for Model Parameter Estimation*. Elsevier, New York, 1987.

[19] D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, May 1996.

[20] O. Michailovich and D. Adam. Robust estimation of ultrasound pulses using outlier-resistant de-noising. *IEEE Transactions on Medical Imaging*, 22(3):368–392, March 2003.

[21] T. F. Chan and C.-K. Wong. Total variation blind deconvolution. *IEEE Transactions Image Processing*, 7(3):370–375, March 1998.

[22] M. A. T. Figueiredo and J. Bioucas Dias. Restoration of Poissonian images using alternating direction optimization. *IEEE transactions on Image Processing*, 19(12):3133–3145, December 2010.

[23] E. Candes and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies. *IEEE Transactions on Information Theory*, 52(12):5406–5425, December 2006.

[24] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1221, 2006.

[25] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[26] E. Candes and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Computational Mathematics*, 6(2):227–254, 2006.

[27] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. In *Princeton University Bulletin*, pages 49–52.

[28] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[29] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[30] J. M. Bioucas-Dias and M. A. T. Figueiredo. Multiplicative noise removal using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(7):1720–1730, 2010.

[31] E. Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. CAM Report 09–31, UCLA, 2009.

[32] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(9):2345–2356, September 2010.

[33] T. Goldstein and S. Osher. The split Bregman method for L1 regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[34] J. Eckstein. *Splitting Methods for Monotone Operators with applications to Parallel Optimization*. PhD thesis, Massachusetts Institute of Technology, 1989.

[35] R. Glowinski and A. Marroco. Sur L'Approximation, par Elements finis d'ordre Un, et la Resolution, par Penalization-Dualité, d'une Classe de Problemes de Dirichlet nonlineares. *Revue Francaise d'Automatique, Informatique et Recherche Opérationelle*, R-2:41–76.

[36] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximations. *Computers and Mathematics with Applications*, 2:17–40.

[37] J. Eckstein and D. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55, 1992.

[38] L. P. Yaroslavsky. *Digital Picture Processing-An introduction*. Springer-Verlag, 1985.

[39] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision*, pages 839–846, Washington, DC, USA, 1998. IEEE Computer Society.

[40] H. L. Lutsep, G. W. Albers, A. DeCrespigny, G. N. Kamat, M. P. Marks, and M. E. Moseley. Clinical utility of diffusion-weighted magnetic resonance imaging in the assessment of ischemic stroke. *Annals of Neurology*, 41:574–580, May 1997.

[41] O. Michailovich and A. Tannenbaum. Blind deconvolution of medical ultrasound images: Parametric inverse filtering approach. *IEEE Transactions on Image Processing*, 16(12):3005–3019, December 2007.

[42] A. Tsai, A. Yezzi, and A. S. Willsky. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Transactions on Image Processing*, 10(8):1169–1186, August 2001.

[43] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Image enhancement and denoising by complex diffusion processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1020–1036, August 2004.

[44] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, December 2006.

[45] G. Kim. PDE-based image restoration: a hybrid model and color image denoising. *IEEE Transactions on Image Processing*, 15(5):1163–1170, May 2006.

[46] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Estimation of optimal PDE-based denoising in the SNR sense. *IEEE Transactions on Image Processing*, 15(8):2269–2280, August 2006.

[47] G. Aubert and J. F. Aujol. A variational approach to removing multiplicative noise. *SIAM Journal of Applied Mathematics*, 68(4):925–946, 2008.

[48] S. G. Chang, Y. Bin, and M. Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546, September 2000.

[49] 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.

[50] F. Abramovich, T. Sapatinas, and B.W. Silverman. Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society B.*, 60:725–749, 1998.

[51] H. A. Chipman, E. D. Kolaczyk, and R. E. Mcculloch. Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association*, 92:1413–1421, December 1997.

[52] R. R. Coifman and D. L. Donoho. Translation-invariant de-noising. In *Wavelets and Statistics*, pages 125–150. Springer-Verlag, 1995.

[53] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[54] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1224, 1995.

[55] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola. From local kernel to nonlocal multiple-model image denoising. *International Journal of Computer Vision*, 86:1–32, 2010.

[56] V. Aurich and J. Weule. Non-linear Gaussian filters performing edge preserving diffusion. In *DAGM-Symposium*, pages 538–545. Springer-Verlag, 1995.

[57] S. M. Smith and J. M. Brady. SUSAN - A new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, May 1997.

[58] T.Q. Pham and L.J. van Vliet. Separable bilateral filtering for fast video preprocessing. In *IEEE International Conference on Multimedia and Expo (ICME)*, July 2005.

[59] W. Jiang, M. L. Baker, Q. Wu, C. Bajaj, and W. Chiu. Applications of a bilateral denoising filter in biological electron microscopy. *Journal of Structural Biology*, 144(1-2):114 – 122, 2003.

[60] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '02, pages 257–266, 2002.

[61] M. Elad. On the origin of the bilateral filter and ways to improve it. *IEEE Transactions on Image Processing*, 11(10):1141 – 1151, October 2002.

[62] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. Technical report, Massachusetts Institute of Technology, 2006.

[63] 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.

[64] 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.

[65] P. Coupé, P. Hellier, C. Kervrann, and C. Barillot. Nonlocal means-based speckle filtering for ultrasound images. *IEEE Transactions on Image Processing*, 18(10):2221–2229, October 2009.

[66] J. V. Manjón, 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.

[67] C. A. Deledalle, F. Tupin, and L. Denis. Poisson NL means: Unsupervised non local means for Poisson noise. In *IEEE International Conference on Image Processing*, September 2010.

[68] C. Kervrann, J. Boulanger, and P. Coupé. Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal. In *Proceedings of the Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, 2007.

[69] T. Teuber and A. Lang. A new similarity measure for nonlocal filtering in the presence of multiplicative noise. *Computational Statistics and Data Analysis*, 56(12):3821–3842, 2012.

[70] S. Haykin. *Blind Deconvolution*. Prentice Hall, Englewood Cliffs, NJ, 1994.

[71] J. A. Cadzow. Blind deconvolution via cumulant extrema. *IEEE Signal Processing Magazine*, 13(3):24–42, May 1996.

[72] A. S. Carasso. Linear and nonlinear image deblurring: A documented study. *SIAM Journal on Numerical Analysis*, 36(6):1659–1689, 1999.

[73] M. Ben-Ezra and S. K. Nayar. Motion-based motion deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):689–698, June 2004.

[74] Y. Wang, W. Yin, and Y. Zhang. A fast algorithm for image deblurring with total variation regularization. Technical Report TR07-10, Rice University, 2010.

[75] R. H. T. Bates. Astronomical speckle imaging. *Physics Reports*, 90(4):203–297, October 1982.

[76] T. J. Schulz. Multiframe blind deconvolution of astronomical images. *Journal of Optical Society of America A*, 10(5):1064–1073, 1993.

[77] J. P. Muller. *Digital Image Processing in Remote Sensing*. Taylor and Francis, 1988.

[78] J. A. Jensen, J. Mathorne, T. Gravesen, and B. Stage. Deconvolution of in vivo ultrasound B-mode images. *Ultrasonic Imaging*, 15(2):122–133, April 1993.

[79] U. R. Abeyratne, A. P. Petropulu, and J. M. Reid. Higher order spectra based deconvolution of ultrasound images. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 42(6):1064–1075, November 1995.

[80] O. Michailovich and D. Adam. A novel approach to the 2-D blind deconvolution problem in medical ultrasound. *IEEE Transactions on Medical Imaging*, 24(1):86–104, January 2005.

[81] L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astronomical Journal*, 79(6):745–754, 1974.

[82] A. K. Katsaggelos. *Digital Image Restoration*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1991.

[83] A. Jain and S. Ranganath. Application of two dimensional spectral estimation in image restoration. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 1113 – 1116, April 1981.

[84] J. Biemond, R. L. Lagendijk, and R. M. Mersereau. Iterative methods for image deblurring. *Proceedings of the IEEE*, 78(5):856–883, May 1990.

[85] T. Taxt. Restoration of medical ultrasound images using two-dimensional homomorphic deconvolution. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 42(4):543–554, July 1995.

[86] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Understanding blind deconvolution algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:2354–2367, 2011.

[87] R. G. Paxman, T. J. Schulz, and J. R. Fienup. Joint estimation of object and aberrations by using phase diversity. *Journal of the Optical Society of America A*, 9(7):1072–1085, July 1992.

[88] J. Markham and J.-A. Conchello. Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur. *Journal of the Optical Society of America A*, 16(10):2377–2391, October 1999.

[89] T. J. Holmes. Blind deconvolution of quantum-limited incoherent imagery: Maximum-likelihood approach. *Journal of the Optical Society of America A*, 9(7):1052–1061, July 1992.

[90] M. Unser. Sampling-50 years after shannon. *Proceedings of the IEEE*, 88(4):569–587, April 2000.

[91] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[92] Y. Tsaig and D. Donoho. Extensions of compressed sensing. *Signal Processing*, 86(3):533–548, 2006.

[93] J. Haupt and R. Nowak. Signal reconstruction from noisy random projections. *IEEE Transactions on Information Theory*, 52(9):4036–4048, 2006.

[94] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[95] R. Prony. Essai expérimental et analytique sur les lois de la dilatabilité des uides élastique et sur celles de la force expansive de la vapeur de leau et de la vapeur de lalkool, à diff erentes températures. *J. École Polytechnique*, 1:24–76, 1795.

[96] B. Logan. *Properties of High Pass Signals*. PhD thesis, Columbia University, 1965.

[97] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, December.

[98] U. Gamper, P. Boesiger, and S. Kozerke. Compressed sensing in dynamic MRI. *Magnetic Resonance in Medicine*, 59(2):365–373, February 2008.

[99] J. Provost. The application of compressed sensing for photo-acoustic tomography. *IEEE Transactions on Medical Imaging*, 28(4):585 – 594, April 2009.

[100] H. Yu and G. Wang. Compressed sensing based interior tomography. *Physics in Medicine and Biology*, 54(9):2791–2805, 2009.

[101] W. Dai, M. A. Sheikh, O. Milenkovic, and R. G. Baraniuk. Compressed sensing DNA Microarrays. *EURASIP Journal on Bioinformatics and Systems Biology*, 2009.

[102] M. Mohtashemi, H. Smith, D. Walburger, F. Sutton, and J. Diggans. Sparse sensing DNA microarray-based biosensor: Is it feasible? In *Sensors Applications Symposium (SAS), 2010 IEEE*, pages 127 –130, February 2010.

[103] T. Lin and F. J. Herrmann. Compressed wavefield extrapolation. *Geophysics*, 72(5):SM77–SM93, September-October 2007.

[104] A. Gholami and H. R. Siahkoohi. Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints. *Geophysical Journal International*, 180(2):871–882, February 2010.

[105] R. M. Willett, M. E. Gehm, and D. J. Brady. Multiscale reconstruction for computational spectral imaging. In *Proceedings of SPIE*, volume 6498, 2007.

[106] J. Bobin, J.-L. Starck, and R. Ottensamer. Compressed sensing in astronomy. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):718–726, October 2008.

[107] 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.

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

[109] P. J. Basser, J. Mattiello, and D. LeBihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, 1994.

[110] 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.

[111] 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.

[112] 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.

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

[114] 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.

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

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

[117] 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.

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

[119] 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.

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

[121] 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.

[122] C. P. Lin, V. J. Wedeen, J. H. Chen, C. Yao, and W. Y. Tseng. Validation of diffusion spectrum magnetic resonance imaging with manganese-enhanced rat optic tracts and ex vivo phantoms. *NeuroImage*, 19(3):482–495, 2003.

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

[124] C. P. Hess, P. Mukherjee, E. T. Han, D. Xu, and D. R. Vigneron. Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. *Magnetic Resonance in Medicine*, 56:104–117, 2006.

[125] 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.

[126] O. Michailovich and Y. Rathi. Fast and accurate reconstruction of HARDI data using compressed sensing. In *Lecture Notes in Computer Science*, volume 6361, pages 606–613, Beijing, China, September 2010. MICCAI.

[127] B. A. Landman, H. Wan, J. A. Bogovic, P.-L. Bazin, and J. L. Prince. Resolution of crossing fibers with constrained compressed sensing using traditional Diffusion Tensor MRI. In *Proceedings of SPIE*, Medical Imaging, San Diego, CA, February 2010.

[128] B. A. Landman, H. Wan, J. A. Bogovic, P. C. M. Van Zijl, P.-L. Bazin, and J. L. Prince. In the pursuit of intra-voxel fiber orientations: Comparison of compressed sensing DTI and Q-ball MRI. In *Proceedings of ISMRM*, Stockholm, Sweden, 2010.

[129] S. Merlet and R. Deriche. Compressed sensing for accelerated EAP recovery in diffusion MRI. In *MICCAI Workshop on Computational Diffusion MRI*, 2010.

[130] M. I. Menzel, K. Khare, K. F. King, X. Tao, C. J. Hardy, and L. Marinelli. Accelerated diffusion spectrum imaging in the human brain using compressed sensing. In *Proceedings of ISMRM*, Stockholm, Sweden, 2010.

[131] N. Lee and M. Singh. Compressed sensing based Diffusion Spectrum Imaging. In *Proceedings of ISMRM*, Stockholm, Sweden, 2010.

[132] O. Michailovich and Y. Rathi. On approximation of orientation distributions by means of spherical ridgelets. *IEEE Transactions on Image Processing*, 19(2):461–477, February 2010.

[133] P. Combettes and V. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM journal on Multiscale Modeling and Simulation*, 4:1168–1200, 2005.

[134] P. L. Combettes and J.-C. Pesquet. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, December 2007.

[135] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[136] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak. Majorization minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 16(12):2980–2991, December 2007.

[137] L. M. Bregman. A relaxation method for finding a common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, (7):200–217, 1967.

[138] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman itrative algorithms for ℓ_1-minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168, 2008.

[139] D. P. Bertsekas, A. N., and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Belmont, MA, 2003.

[140] D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part 1 – theory. *IEEE Transactions on Medical Imaging*, 1(2):81–94, October 1982.

[141] J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes Rendus de l' Acadḿie dex Sciences de Paris*, 255:2897–2899, 1962.

[142] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493, July 2009.

[143] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

[144] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[145] E. Candes and J. Romberg. Sparsity and incoherence in compressive sampling. *Comptes Rendus Mathematique*, 23(3):969–985, 2007.

[146] E. Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9-10):589–592, 2008.

[147] D. Manolakis, C. Siracusa, and G. Shaw. Hyperspectral subpixel target detection using the linear mixing model. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1392–1409, July 2001.

[148] C.-I. Chang and D. C. Heinz. Constrained subpixel target detection for remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1144–1159, May 2000.

[149] M. B. Lopes, J. C. Wolff, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Near-infrared hyperspectral unmixing based on a minimum volume criterion for fast and accurate chemometric characterization of counterfeit tablets. *Analytical Chemistry*, 82:1462–1469, 2010.

[150] S. M. Chai, A. Gentile, W. E. Lugo-Beauchamp, J. L. Cruz-Rivera, and D. S. Wills. Hyper-spectral image processing applications on the SIMD pixel processor for the digital battlefield. In *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Method and Applications*, pages 130–138, 1999.

[151] D. S. Tuch, R. M. Weisskoff, J. W. Belliveau, and V. J. Wedeen. High angular resolution diffusion imaging of the human brain. In *Proceedings of the 7th Annual Meeting of ISMRM*, page 321, 1999.

[152] D. Li, K. Wong, Y. H. Hu, and A. Sayeed. Detection, classification, and tracking of targets. *IEEE Signal Processing Magazine*, 19(2):17–29, March 2002.

[153] A. Arora et al. A line in the sand: a wireless sensor network for target detection, classification, and tracking. *Computer Networks*, 46(5):605–634, 2004.

[154] B. D. Geelhood, J. H. Ely, R. R. Hansen, R. T. Kouzes, J. E. Schweppe, and R. A. Warner. Overview of portal monitoring at border crossings. In *2003 IEEE Nuclear Science Symposium Conference Record*, volume 1, pages 513–517, October 2003.

[155] J. Agre and L. Clare. An integrated architecture for cooperative sensing networks. *Computer*, 33(5):106–108, May 2000.

[156] K. Bult et al. Low power systems for wireless microsensors. In *UCLA Electrical Engineering Department Los Angeles, CA and Rockwell Science Center Thousand Oaks California*, 2002.

[157] A. Chandrakasan, R. Amirtharajah, S. Cho, J. Goodman, G. Konduri, J. Kulik, W. Rabiner, and A. Wang. Design considerations for distributed microsensor systems. In *Proceedings of the IEEE Custom Integrated Circuits Conference (CICC)*, pages 279–286, 1999.

[158] M. Dong, K. Yung, and W. Kaiser. Low power signal processing architectures for network microsensors. In *Proceedings of International Symposium on Low Power Electronics and Design*, pages 173–177, August 1997.

[159] S. M. Brennan, A. M. Mielke, D. C. Torney, and A. B. Maccabe. Radiation detection with distributed sensor networks. *Computer*, 37(8):57–59, 2004.

[160] M. Mahfuz and K. Ahmed. A review of micro-nano-scale wireless sensor networks for environmental protection: Prospects and challenges. *Science and Technology of Advanced Materials*, 6(3–4):302–306, May 2005.

[161] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, November 2006.

[162] P. Blomberg and T. Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, March 1998.

[163] B. Goldluecke, E. Strekalovskiy, and D. Cremers. The natural vectorial total variation which arises from geometric measure theory. *SIAM Journal on Imaging Sciences*, 5(2):537 –563, 2012.

[164] M. S. Asif and J. K. Romberg. Dynamic updating for ℓ_1 minimization. *Journal of Selected Topics in Signal Processing*, 4(2):421–434, 2010.

[165] S. Mori. *Introduction to diffusion tensor imaging*. Elsevier, 2007.

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

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

[168] M. C. Roggemann and B. M. Welsh. *Imaging through turbulence*. CRC press, 1996.

[169] B. Angelsen. *Ultrasound imaging: waves, signals, and signal processing*. Emantec, Trondhejm, Norway, 2000.

[170] O. Michailovich and A. Tannenbaum. De-speckling of medical ultrasound images. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 53(1):64–78, January 2006.

[171] H. L. Taylor, S. C. Banks, and J. F. McCoy. Deconvolution with ℓ_1 norm. *Geophysics*, 44(1):39–52, January 1979.

[172] 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.

[173] 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 (MICCAI), Part II*, pages 171–179, 2008.

[174] Y. Matsushita and S. Lin. A probabilistic intensity similarity measure based on noise distributions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

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

[176] M. Seeger. *Covariance Kernels from Bayesian Generative Models*, pages 905–912. 2002.

[177] J. Polzehl and V. Spokoiny. Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields*, 135(3):335–362, 2006.

[178] J. W. Goodman. Some fundamental properties of speckle. *Journal of the Optical Society of America*, 66(11):1145–1150, November 1976.

[179] M. Tur, K. C. Chin, and J. W. Goodman. When is speckle noise multiplicative. *Applied Optics*, 21(7):1157–1159, 1982.

[180] J. W. Goodman. *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers, 2007.

[181] L. Rudin, P. Lions, and S. Osher. Multiplicative denoising and deblurring: Theory and algorithms. *Geometric Level Set Methods in Imaging, Vision, and Graphics*, pages 103–119, 2003.

[182] T. C. Aysal and K. E. Barner. Rayleigh-maximum-likelihood filtering for speckle reduction of ultrasound images. *IEEE Transactions on Medical Imaging*, 26(5):712–727, 2007.

[183] B. B. Saevarsson, J. R. Sveinsson, and J. A. Benediktsson. Combined wavelet and curvelet denoising of SAR images. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, pages 4235–4238, 2004.

[184] G. Steidl and T. Teuber. Removing multiplicative noise by Douglas-Rachford splitting methods. *Journal of Mathematical Imaging and Vision*, 36(2):168–184, 2010.

[185] L. Denis, F. Tupin, J. Darbon, and M. Sigelle. SAR image regularization with fast approximate discrete minimization. *IEEE Transactions on Image Processing*, 18(7):1588–1600, 2009.

[186] S. Durand, J. Fadili, and M. Nikolova. Multiplicative noise removal using L1 fidelity on frame coefficients. *Journal of Mathematical Imaging and Vision*, 36(3):201–226, 2010.

[187] C. Chinrungrueng and A. Suvichakorn. Fast edge-preserving noise reduction for ultrasound images. *IEEE Transactions on Nuclear Science*, 48(3):849–854, 2001.

[188] E. E. Kuruoglu. Modeling SAR images with a generalization of the Rayleigh distribution. *IEEE Transactions on Image Processing*, 13(4):527–533, 2004.

[189] T. Eltoft. Modeling the amplitude statistics of ultrasonic images. *IEEE Transactions on Medical Imaging*, 25(2):229–240, 2006.

[190] Z. Zhao and L. Cheng. Variational image segmentation in presence of multiplicative Gamma noise. *Electronic letters*, 47(16):918–919, 2011.

[191] L.-L. Huang, L. Xiao, and Z.-H. Wei. Multiplicative noise removal via a novel variational model. *EURASIP Journal on Image and Video Processing*, 2010.

[192] M. Yavuz and J. A. Fessler. Statistical image reconstruction methods for randomsprecorrected PET scans. *Medical Image Analysis*, 2(4):369–378, 1998.

[193] S.-J. Lee, A. Rangarajan, and G. Gindi. Bayesian image reconstruction in SPECT using higher order mechanical models as priors. *IEEE Transactions on Medical Imaging*, 14(4):669–680, December 1995.

[194] R. A. Boie and I. J. Cox. An analysis of camera noise. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 14(6):671–674, June 1992.

[195] G. E. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, March 1994.

[196] H. Bradt. *Astronomy methods: A physical approach to astronomical observations*. Cambridge University Press, 2004.

[197] E. Bratsolis and M. Sigelle. A spatial regularization method preserving local photometry for Richardson-Lucy restoration. *Astronomy and Astrophysics*, 375(3):1120–1128, 2001.

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

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

[200] J. V. Manjón, 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.

[201] R. D. Nowak. Wavelet-based Rician noise removal for magnetic resonance imaging. *IEEE Transactions on Image Processing*, 8(10):1408–1419, October 1999.

[202] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[203] M. L. Glasser and E. Montaldi. Some integrals involving Bessel functions. *Journal of Mathematical Analysis and Applications*, 183(1):577–590, May 1994.